

Tradução de David Ragazzi Rocha

Índice

Prólogo

1. Inteligência Artificial

2. Redes Neurais

3. O Cérebro Humano

4. A Memória

5. Um Novo Modelo para a Inteligência

6. Como Funciona o Córtex

7. Consciência e Criatividade

8. O Futuro da Inteligência

Epílogo

Apêndice: Predições Verificáveis

Bibliografia

Agradecimentos

Sobre a Inteligência

Prólogo

Duas paixões animam este livro e minha vida.

Durante vinte e cinco anos tenho sido um entusiasta dos computadores portáteis. No mundo de alta tecnologia do Vale do Silício me conhecem por ter posto em marcha duas empresas, Palm Computing e Handspring, e por ser o criador de muitos computadores de mão e telefones móveis como o PalmPilot e o Treo.

Mas tenho uma segunda paixão que precede a meu interesse pelos computadores e que a considero mais importante: eu sou louco por cérebros. Quero entender como funcionam não só de uma perspectiva filosófica, mas também do modo detalhado da engenharia, analisando seus elementos básicos. Meu desejo não se limita a compreender o que é a inteligência e como atua o cérebro, mas pretendo aprender a construir máquinas que funcionem da mesma maneira. Quero construir máquinas realmente inteligentes.

A questão da inteligência é a última grande fronteira terrestre da ciência. A maioria dos temas científicos importantes trata do muito pequeno, do muito grande ou de acontecimentos que ocorreram há bilhões de anos. Mas toda pessoa tem um cérebro. Você é seu cérebro. Se você quer compreender por que se sente como se sente, como percebe o mundo, por que comete erros, como é capaz de ser criativo, por que a música e a arte lhe servem de inspiração, em definitivo, o que é ser humano, você precisa entender o cérebro. Além disso, uma teoria satisfatória sobre a inteligência e a função cerebral proporcionará amplos benefícios sociais, que não se limitarão a nos ajudar a curar as doenças relacionadas com o cérebro. Seremos capazes de construir máquinas realmente inteligentes, ainda que não se parecerão de forma alguma aos robôs da ficção popular e da fantasia da ciência da computação. Preferencialmente, as máquinas inteligentes surgirão de um novo conjunto de princípios sobre a natureza da inteligência. As ditas máquinas nos ajudarão a acelerar nosso conhecimento do mundo, nos ajudarão na exploração do Universo e farão com que o mundo seja mais seguro. E de passagem se criará uma enorme indústria.

Por sorte, vivemos em uma época na qual o problema de entender a inteligência pode ser resolvido. Nossa geração tem acesso a uma montanha de dados sobre o cérebro reunidos durante centenas de anos, e o ritmo com que seguimos os reunindo está se acelerando. Só nos Estados Unidos há milhares de neurocientistas. No entanto, não contamos com teorias produtivas sobre o que é a inteligência ou como funciona o cérebro em seu conjunto. A maioria dos neurobiólogos não pensa muito sobre teorias gerais do cérebro porque estão concentrados em realizar experimentos para reunir mais dados sobre os muitos subsistemas cerebrais. E ainda que legiões de programadores de computador têm tentado tornar inteligentes os computadores, têm fracassado. E acho que

continuarão fracassando enquanto continuarem ignorando as diferenças entre computadores e cérebros.

Portanto, em que consiste a inteligência da qual possui o cérebro, mas não os computadores? Por que uma criança de seis anos pode saltar com desenvoltura de rocha em rocha em um ribeirão, enquanto os robôs mais avançados da nossa época são zumbis desajeitados? Por que crianças de três anos estão a caminho de dominar a linguagem, enquanto os computadores parecem incapazes de fazer isto apesar de meio século de esforços dos programadores? Por que podemos discernir o que é um gato e não um cão em uma fração de segundo, enquanto um supercomputador não é capaz de estabelecer a distinção? São grandes mistérios que esperam uma resposta. Temos pistas em abundância; agora o que precisamos são de algumas percepções decisivas.

Talvez se pergunte por que um projetista de computadores escreve um livro sobre cérebros. Ou expressado de outro modo: se gosto dos cérebros, por que não desenvolvi minha carreira em ciência cerebral ou em inteligência artificial. A resposta é que tentei várias vezes, mas me neguei a estudar o problema da inteligência como tinham feito outros antes de mim. Achava que o melhor modo de resolver este problema era empregar a biologia detalhada do cérebro como um delimitador e uma pauta, porém concebendo a inteligência como um problema computacional, em uma postura que o situasse em um ponto entre a biologia e a ciência da computação. Muitos biólogos tendem a recusar ou ignorar a ideia de pensar no cérebro em termos computacionais, e os cientistas da computação costumam achar que não têm nada que aprender da biologia. Além disso, o mundo da ciência aceita menos o risco do que o empresarial. Nas empresas tecnológicas, uma pessoa que persegue uma nova ideia com uma proposta razoável pode ascender na sua carreira seja qual for o sucesso que atinja a dita ideia. Muitos empresários conseguiram triunfar após vários fracassos anteriores. Mas no mundo acadêmico, um par de anos dedicados a uma nova ideia que não dá bons resultados pode arruinar para sempre uma carreira que se inicia. Portanto, dediquei-me às duas paixões da minha vida de forma simultânea, achando que o sucesso na indústria me ajudaria a me dedicar à ciência que eu desejava; além disso, eu precisava aprender como ter afinidade com a mudança no mundo e como vender novas ideias, e tudo isso eu esperava obter trabalhando no Vale do Silício.

Em agosto de 2002 pus em marcha um centro de pesquisas, o Redwood Neuroscience Institute (RNI), dedicado à teoria sobre o cérebro. Existem muitos centros de neurociência no mundo, mas nenhum outro se dedica ao entendimento teórico geral do neocórtex, a parte do cérebro humano responsável pela inteligência. Isto é a única coisa que estudamos no RNI, que em muitos aspectos se assemelha a uma empresa que se inicia. Perseguimos um sonho que algumas

peças consideram inalcançável, mas temos a sorte de sermos um grande grupo, e nossos esforços estão começando a dar frutos.

* * *

O conteúdo deste livro é ambicioso. Descreve uma teoria exaustiva de como funciona o cérebro. Detalha o que é a inteligência e como a cria seu cérebro. A teoria que apresento não é completamente nova. Muitas das ideias particulares que você está a ponto de ler já existiam de uma forma ou de outra, mas não juntas de um modo coerente. Isto é algo que era de se esperar. Dizem que as "novas ideias" costumam ser velhas ideias reordenadas e reinterpretadas, o que sem dúvida se torna aplicável à teoria que se propõe neste texto; mas o reordenamento e a interpretação podem significar uma diferença abismal, a diferença entre um cúmulo de detalhes e uma teoria satisfatória. Espero que ela lhe cause a mesma impressão que provoca em muita gente. A reação típica que escuto é: "Faz sentido. Não tinha pensado na inteligência dessa maneira, mas agora que você me tem descrito posso ver como tudo se encaixa". Com este conhecimento a maioria das pessoas começa a ver a si mesmas de uma forma um pouco diferente. Você começa a observar seu comportamento dizendo: "Compreendo o que acaba de acontecer na minha cabeça". Espero que quando você terminar de ler este livro você já tenha uma nova percepção de por que você pensa como pensa e porque se comporta como se comporta. Também espero que alguns leitores se sintam inspirados para dirigir suas carreiras à construção de máquinas inteligentes baseadas nos princípios esboçados nestas páginas.

Costumo me referir a esta teoria e a minha proposta de estudo da inteligência como "inteligência real" para distinguir da "inteligência artificial". Os cientistas da inteligência artificial tentaram programar computadores para que atuassem como seres humanos sem solucionar primeiro o que é a inteligência e o que significa compreender. Deixaram de lado a parte mais importante para construir máquinas inteligentes: a inteligência. A "inteligência real" estabelece a premissa de que antes de tentarmos construir máquinas inteligentes devemos compreender primeiro como pensa o cérebro, e não há nada artificial nisso. Só então poderemos nos perguntar como construir máquinas inteligentes.

O livro inicia-se com alguns antecedentes de por que fracassaram as tentativas anteriores de compreender a inteligência e construir máquinas inteligentes. Depois apresento e desenvolvo a ideia central da teoria, o que chamo de modelo de memória-predição. No capítulo 6 detalho como o cérebro físico põe em prática o modelo de memória-predição; em outras palavras, como funciona realmente o cérebro. A seguir explico as repercussões sociais e outras da teoria no que para muitos leitores talvez constitua a epígrafe que mais reflexões suscita. O livro conclui com um exame das máquinas inteligentes: como podemos as construir e como será o futuro. Espero que isto lhe pareça fascinante. Estas são algumas das perguntas das quais cobriremos ao longo do texto:

Podem ser inteligentes os computadores?

Durante décadas os cientistas do campo da inteligência artificial têm declarado que os computadores serão inteligentes quando contarem com a potência necessária. Não acredito, e explicarei por que. Os cérebros e os computadores fazem coisas radicalmente diferentes.

Não se acreditava que as redes neurais conduziriam às máquinas inteligentes?

Sabemos que o cérebro está constituído por uma rede de neurônios, mas se não entendemos primeiro o que é que ela faz, algumas redes neurais simples não obterão melhores resultados em criar máquinas inteligentes do que os programas de computador.

Por que tem sido tão difícil entender como funciona o cérebro?

A maioria dos cientistas afirma que devido à complexidade do cérebro demoraremos muito tempo em compreendê-lo. Não estou de acordo. A complexidade é um sinónimo de confusão, não uma causa. O que sustento é que algumas hipóteses intuitivas, porém inexatas, nos induziram ao erro. A maior foi achar que a inteligência é definida pelo comportamento inteligente.

O que é a inteligência se ela não se define pelo comportamento?

O cérebro emprega enormes quantidades de memória para criar um modelo do mundo. Tudo o que você conhece e tem aprendido se armazena neste modelo. O cérebro usa o dito modelo baseado na memória para efetuar previsões contínuas sobre acontecimentos futuros. A capacidade para efetuar previsões sobre o futuro constitui o núcleo da inteligência. Descreverei a capacidade preditiva do cérebro a fundo, pois esta é a ideia central deste livro.

Como funciona o cérebro?

A base da inteligência é o neocórtex. Ainda que possua um grande número de capacidades e uma flexibilidade enorme, o neocórtex é surpreendentemente regular em todos os seus detalhes estruturais. Suas partes distintas, sejam elas responsáveis pela visão, audição, tato ou linguagem, funcionam segundo os mesmos princípios. A chave para entender o neocórtex é compreender estes princípios comuns e, em particular, sua estrutura hierárquica. Examinaremos o neocórtex em detalhes suficientes para mostrar como sua estrutura capta a configuração do mundo. Esta explicação será a parte mais técnica do livro, porém os

leitores interessados serão capazes de compreendê-la ainda que não sejam cientistas.

Quais são as repercussões desta teoria?

A teoria do cérebro pode ajudar a explicar coisas tais como nossa maneira de sermos criativos, por que temos consciência de algo, por que mostramos preconceitos, como aprendemos e por que “cachorros velhos” costumam a aprender “novos truques”. Analisarei vários destes temas. Em linhas gerais, esta teoria proporciona-nos uma percepção do que somos e de por que fazemos o que fazemos.

Podemos construir máquinas inteligentes? O que elas farão?

Sim. Podemos e faremos. Considero que as capacidades das ditas máquinas evoluirão rapidamente em algumas décadas e em direções interessantes. Algumas pessoas temem que as máquinas inteligentes possam se tornar perigosas para a humanidade, porém não compartilho essa ideia. Os robôs não vão nos invadir. Por exemplo, será bem mais fácil construir máquinas que ultrapassem nossas faculdades em pensamento elevado de âmbitos como a física e a matemática do que construir algo parecido aos robôs falantes que vemos na ficção popular. Explorarei as incríveis direções nas quais é provável que avance esta tecnologia.

Minha meta é explicar esta nova teoria da inteligência e do funcionamento do cérebro de um modo que qualquer pessoa seja capaz de entender. Uma boa teoria deve se tornar fácil de compreender, e não se ocultar em jargão ou em um argumento confuso. Começarei com um modelo básico e depois irei acrescentando detalhes à medida que avançarmos. Alguns raciocinarão atendendo à lógica; outros envolverão aspectos particulares do sistema de circuitos cerebral. Sem dúvida, alguns dos detalhes que proponho serão errôneos, o que sempre acontece em qualquer campo da ciência. Demorará anos para ser desenvolvida uma teoria plenamente acabada, mas isso não diminui a força da ideia central.

* * *

Quando comecei a me interessar pelos cérebros há muitos anos, fui à biblioteca local para buscar um bom livro que explicasse o funcionamento cerebral. Desde adolescente tinha-me acostumado a ser capaz de encontrar livros bem escritos que explicavam quase qualquer tema que me interessasse. Tinha os sobre a teoria da relatividade, buracos negros, matemática, sobre qualquer coisa que me fascinara até então. No entanto, minha busca de um livro satisfatório sobre o cérebro foi em vão. Acabei me dando conta de que ninguém tinha ideia de como ele funcionava na realidade. Nem sequer existiam teorias ruins ou não validadas; simplesmente não existiam, o que era algo pouco habitual. Por exemplo, até então ninguém

sabia como tinham morrido os dinossauros, mas existia uma multidão de teorias, e de todas você podia aprender algo. Mas não era assim no caso dos cérebros. A princípio custei a acreditar. Preocupava-me o fato de que não sabíamos como funcionava este órgão crucial. Enquanto estudava o que já sabíamos sobre os cérebros, cheguei a pensar que deveria existir uma explicação simples. O cérebro não era mágico, e não me parecia que as respostas pudessem ser tão complexas. O matemático Paul Erdos achava que as provas matemáticas mais simples já existiam em um livro etéreo e que a tarefa do matemático era as encontrar, "ler o livro". Do mesmo modo, parecia-me que a explicação da inteligência estava "lá fora". Podia apalpá-lo. Eu queria ler o livro.

Durante os últimos vinte e cinco anos tenho tido uma visão desse livro pequeno e claro sobre o cérebro. Era como uma cenoura que me mantinha motivado durante todo esse tempo. Esta visão deu forma ao livro que você tem agora em suas mãos. Nunca gostei de complexidade, nem na ciência nem na tecnologia. Você pode ver isso refletido nos produtos que tenho projetado, que costumam ser conhecidos por sua facilidade de uso. As coisas mais convincentes são simples. Portanto, este livro propõe uma teoria simples e direta sobre a inteligência. Espero que você goste.

1

Inteligência Artificial

Quando em junho de 1979 obtive o diploma em engenharia elétrica pela Cornell, não tinha nenhum plano importante para minha vida. Comecei a trabalhar como engenheiro no novo campus de Intel em Portland (Oregón). A indústria de microcomputadores estava começando, e a Intel achava-se no seu centro. Meu trabalho consistia em analisar e resolver os problemas encontrados pelos demais engenheiros que realizavam trabalho de campo com nosso produto principal, os computadores de placa única. (Até então, a recente invenção do microprocessador por parte da Intel tinha tornado possível que se pusesse um computador inteiro em uma única placa de circuito integrado.) Eu publicava um boletim de notícias, realizava algumas viagens e tive a oportunidade de conhecer clientes. Era jovem e tudo ia muito bem, ainda que tivesse saudades da minha namorada da universidade, que tinha aceitado um trabalho em Cincinnati.

Em uns meses depois me topei com algo que iria mudar a direção da minha vida. Esse algo foi o recém publicado número de setembro da *Scientific American*, que se dedicava completamente ao cérebro. Ele reavivou meu interesse de infância pelos cérebros e me pareceu fascinante. Por ele me inteirei da organização, desenvolvimento e química do cérebro, dos mecanismos neurais da visão, do movimento e outras especializações, e da base biológica dos transtornos mentais. Foi um dos melhores números da *Scientific American* que já se publicou. Vários neurocientistas com que tenho falado me confirmaram que ela desempenhou um papel significativo na escolha da sua carreira, igual o que aconteceu a mim.

O artigo final, "Thinking About the Brain", foi escrito por Francis Crick, codescobridor da estrutura do DNA, que até então tinha dirigido seu talento ao estudo do cérebro. Crick sustentava que, apesar da constante acumulação de conhecimento detalhado sobre o cérebro, continuava sendo um profundo mistério seu funcionamento. Os cientistas não costumam escrever sobre o que desconhecem, mas Crick não se importava. Ele era como o menino apontando para o imperador sem roupa. Segundo ele, a neurociência consistia em uma montanha de dados sem uma teoria. Suas palavras exatas eram: "O que chama a atenção é a falta de um modelo amplo de ideias". A meu entender, era o educado modo britânico de dizer: "Não temos nem uma pista de como funciona essa coisa". Era verdade até então e continua sendo na atualidade.

As palavras de Crick foram como uma chamada para mim. O desejo de toda a minha vida de compreender os cérebros e construir máquinas inteligentes ganhou vida. Ainda que eu mal tivesse saído da universidade, decidi mudar de carreira. Iria estudar os cérebros, não só para entender como funcionam, mas também para empregar esse conhecimento como base para novas tecnologias, como construir máquinas inteligentes. Demoraria algum tempo para eu pôr este plano em prática.

Na primavera de 1980 transferi-me para a sede da Intel em Boston para reunir-me com minha futura esposa, que iria começar um curso de pós-graduação. Ocupei uma função que envolvia ensinar os clientes e os empregados a projetar sistemas baseados em microprocessadores. Porém eu tinha a visão focada em uma meta diferente: eu estava tentando encontrar alguma forma de trabalhar na teoria do cérebro. O engenheiro que existia em mim se dava conta de que uma vez que compreendêssemos como funcionavam os cérebros poderíamos os construir, e o modo natural de construir cérebros artificiais era em silício. Eu trabalhava para a companhia que inventou o chip de memória e o microprocessador de silício, de modo que talvez fosse possível convencer a Intel que me permitisse dedicar parte do meu tempo para pensar sobre a inteligência e como projetar chips de memória semelhantes ao cérebro. Escrevi uma carta ao diretor da Intel, Gordon Moore, que podia se resumir assim:

Estimado doutor Moore:

Proponho que iniciemos um grupo de pesquisas que se dedique a compreender como funciona o cérebro. Pode começar com uma pessoa, eu, e daí por diante. Tenho plena confiança de que somos capazes de o entender. Será um grande negócio algum dia.

—Jeff Hawkins

Moore me pôs em contato com o cientista chefe da Intel, Ted Hoff. Peguei um avião para Califórnia para reunir-me com ele e expor minha proposta de estudar o cérebro. Hoff era famoso por duas coisas. A primeira —da qual eu estava bem inteirado—, era seu trabalho no projeto do primeiro microprocessador. A segunda —da qual eu não conhecia até aquele momento— era seu trabalho na incipiente teoria das redes neurais. Ele tinha experiência com neurônios artificiais e algumas das coisas que você podia fazer com eles. Eu não estava preparado para isto. Depois de escutar minha proposta, ele disse que não lhe parecia possível entender como funciona o cérebro em um futuro previsível e, portanto, não fazia sentido a Intel me apoiar. Hoff estava certo, pois só agora, vinte e cinco anos mais tarde, que estamos começando a efetuar um avanço significativo no entendimento dos cérebros. A oportunidade é tudo nos negócios. Porém, naquele momento senti uma grande desilusão.

Tendo a buscar o caminho com menores obstáculos para atingir minhas metas. Trabalhar com cérebros na Intel teria sido a transição mais simples. Uma vez que foi eliminada essa opção, indaguei sobre a alternativa seguinte. Decidi me apresentar para um curso de pós-graduação no Massachusetts Institute of Technology, que era famoso por suas pesquisas sobre a inteligência artificial e era bem localizado. Parecia encaixar bem. Eu possuía uma ampla formação em ciência da computação —“bom”—. Desejava construir máquinas inteligentes —“bom”—. Queria estudar primeiro os cérebros para ver como funcionam —“epa, isso é um

problema”—. Esta última meta, querer compreender como funcionavam os cérebros, constituía um problema aos olhos dos cientistas do laboratório de inteligência artificial do MIT.

Foi como topar-me contra um muro de concreto. O MIT era a nave-mãe da inteligência artificial. Na época em que solicitei a matrícula no MIT, ele albergava dezenas de pessoas brilhantes cativadas com a ideia de programar computadores para produzir um comportamento inteligente. Para estes cientistas, a visão, a linguagem, a robótica e a matemática não eram mais do que problemas de programação. Os computadores podiam fazer qualquer coisa que um cérebro fizesse e mais, de modo que, por que restringir o pensamento com a desordem biológica do computador da Natureza? Estudar os cérebros limitaria a reflexão. Achavam que era melhor estudar os últimos limites da informática, cuja máxima expressão era os computadores digitais. Seu santo gral era criar programas de computador que primeiro igualassem, e depois, ultrapassassem as faculdades humanas. Adotavam a proposta de que o fim justifica os meios; não lhes interessava como funcionavam os cérebros reais. Alguns se orgulhavam de ignorar a neurobiologia.

Pareceu-me uma forma equivocada de abordar o problema. Intuí que a proposta da inteligência artificial não só fracassaria em criar programas que fizessem o que sabem fazer os humanos, como também não nos ensinaria o que era a inteligência. Os computadores e os cérebros são constituídos segundo princípios completamente diferentes. Um é programado; o outro aprende por si mesmo. Um tem que trabalhar perfeitamente; o outro é flexível por natureza e tolera falhas. Um conta com um processador central; o outro carece de controle centralizado. A lista de diferenças é enorme. A principal razão pela qual eu pensava que os computadores não seriam inteligentes era por que eu compreendia como eles funcionavam até o nível da física dos transistores, e este conhecimento me contribuía na sensação intuitiva de que os cérebros e os computadores eram essencialmente diferentes. Eu não podia provar isto, mas sabia na medida em que se pode saber algo por intuição. No fim das contas, pensei, a inteligência artificial pode levar à invenção de produtos úteis, mas não levará a construir máquinas realmente inteligentes.

Em contraste, eu queria compreender a inteligência e percepção reais, estudar a psicologia e a anatomia do cérebro, enfrentar o desafio de Francis Crick e apresentar um amplo modelo do funcionamento cerebral. Em particular, tinha focada a visão no neocórtex, a parte do cérebro dos mamíferos cujo desenvolvimento era mais recente e onde morava a inteligência. Uma vez que compreendêssemos como funcionava o neocórtex, poderíamos começar a construção de máquinas inteligentes, mas não antes.

Infelizmente, os professores e alunos que conheci no MIT não compartilhavam meus interesses. Não achavam que era preciso estudar cérebros reais para

compreender a inteligência e construir máquinas inteligentes. Assim me disseram. Em 1981 a universidade recusou minha solicitação de matrícula.

* * *

Muitas pessoas creem na atualidade que a inteligência artificial está sã e salva, e que só esperam que os computadores tenham a potência suficiente para tornar realidade suas muitas promessas. Quando os computadores tiverem suficiente memória e potência de processamento —prossegue esta crença—, os programadores da inteligência artificial serão capazes de criar máquinas inteligentes. Não estou de acordo. A inteligência artificial apresenta uma falha fundamental: não aborda de forma adequada o que é a inteligência ou o que significa entender algo. Uma breve olhada na sua história e nos princípios sobre os quais ela se edificou ajudará a explicar como a disciplina se desviou do seu rumo.

A proposta da inteligência artificial nasceu com o computador digital. Uma figura chave do início deste movimento foi o matemático inglês Alan Turing, um dos inventores da ideia do computador de uso geral. Seu golpe mestre foi demonstrar formalmente o conceito de computação universal; isto é, que todos os computadores são equivalentes no básico, apesar dos detalhes da sua construção. Como parte da sua demonstração, ele ilustrou sobre uma máquina imaginária com três partes essenciais: uma caixa de processamento, uma fita de papel e um aparelho que lia e escrevia marcas na fita enquanto ela se movia de um lado a outro. A fita servia para guardar informação, como o famoso código computacional de uns e zeros (isso foi antes da invenção dos chips de memória ou da unidade de disco, de modo que Turing criou a fita de papel para o armazenamento). A caixa, que na atualidade chamamos de unidade de processamento central (CPU), segue um conjunto de regras fixas para ler e editar a informação da fita. Turing demonstrou matematicamente que se você escolher o conjunto de regras preciso para a CPU e lhe for proporcionada uma fita infinitamente longa com a qual trabalhar ela poderá realizar qualquer conjunto de operações do Universo. Seria uma das muitas máquinas equivalentes que agora chamamos de Máquinas de Turing Universais. Se o problema é calcular raízes quadradas ou trajetórias de balas, jogar um jogo, editar fotos ou efetuar transações bancárias, por debaixo disso não há mais do que uns e zeros, e qualquer Máquina de Turing poderia ser programada para efetuar a tarefa. Processamento de informação é processamento de informação que é processamento de informação... Todos os computadores digitais são equivalentes em sua lógica.

A conclusão de Turing era indiscutivelmente verdadeira e muito proveitosa. A revolução da informática e todos seus produtos se basearam nela. Depois, Turing passou à questão de como construir uma máquina inteligente. Ele sentia que os computadores podiam ser inteligentes, mas ele não queria entrar em discussões sobre se era possível ou não. Ele também não se considerava capaz de definir a *inteligência* formalmente, de modo que nem tentou. Ao invés disso, ele propôs

uma prova de existência para a inteligência, o famoso Teste de Turing: se um computador pudesse enganar a um interlocutor humano e conseguisse fazer com que este pensasse que ele também é uma pessoa, o computador teria que ser inteligente por definição. Deste modo, com seu teste como vara de medir e sua máquina como meio, Turing ajudou a lançar a disciplina da inteligência artificial. Seu dogma central: o cérebro não é mais do que outro tipo de computador. Carece de importância o modo como você projeta um sistema artificialmente inteligente; ele só tem que produzir comportamento semelhante ao humano.

Os defensores da inteligência artificial viam paralelos entre os computadores e o pensamento. Diziam: “Vejam, as proezas mais impressionantes da inteligência humana envolvem, sem dúvida, a manipulação de símbolos abstratos, e isso é o que fazem também os computadores. O que fazemos quando falamos ou escutamos? Manipulamos símbolos mentais chamados palavras, utilizando regras gramaticais bem definidas. O que fazemos quando jogamos xadrez? Empregamos símbolos mentais que representam as propriedades e a situação das diversas peças. O que fazemos quando vemos? Usamos símbolos mentais para representar objetos, suas posições, seus nomes e outras propriedades. Sem dúvida, a gente faz tudo isto com os cérebros e não com o tipo de computadores que construímos, mas Turing demonstrou que não era importante como eram executados ou manipulados os símbolos. Você pode fazer isto com uma reunião de dentes e engrenagens, com um sistema de interruptores elétricos ou com a rede de neurônios do cérebro, desde que o meio empregado possa realizar a equivalência funcional de uma Máquina de Turing Universal”.

Esta hipótese foi reforçada por um influente artigo científico publicado em 1943 pelo neuropsicólogo Warren McCulloch e pelo matemático Walter Pitts. Elas descreviam como os neurônios podiam realizar funções digitais; isto é, como cabia a possibilidade de que as células nervosas pudessem reproduzir a lógica formal que constitui o núcleo dos computadores. A ideia era de que os neurônios pudessem agir como o que os engenheiros chamam de portas lógicas. Estas executam operações lógicas simples como E, NÃO e OU. Os chips dos computadores são compostos por milhões de portas lógicas, todas interligadas em circuitos precisos e complexos. Uma CPU não é mais do que um agrupamento de portas lógicas.

McCulloch e Pitts assinalaram que os neurônios também podiam se ligar de formas precisas para realizar funções lógicas. Como os neurônios reúnem contribuições uns dos outros e as processam para decidir se lançam um resultado, era concebível que eles pudessem ser portas lógicas vivas. Deste modo, deduziam, cabia a possibilidade de que o cérebro fosse constituído com portas E, portas OU e outros elementos lógicos criados com neurônios, em analogia direta com a configuração dos circuitos eletrônicos digitais. Não está claro se McCulloch e Pitts acreditavam para valer de que o cérebro funcionava desta forma; só diziam que

era possível. E desde a lógica esta visão dos neurônios é possível. Em teoria, os neurônios podem executar funções digitais. No entanto, ninguém se preocupou em se perguntar se esta era a forma real em que se ligavam os neurônios no cérebro. Tomaram-no como demonstração, sem ter em conta a falta de provas biológicas de que os cérebros não eram mais do que outro tipo de computador.

Também convém assinalar que a filosofia da inteligência artificial foi respaldada pela tendência que dominou a psicologia durante a primeira metade do século XX, chamada comportamentismo. Os comportamentistas achavam que não era possível saber o que acontece dentro do cérebro, o que chamavam de a caixa preta impenetrável. Mas podia-se observar e medir o ambiente de um animal e seus comportamentos, o que ele sente e o que ele faz. Reconheciam que o cérebro continha mecanismos de reflexos que podiam ser usados para condicionar um animal a fim de que ele adotasse novos comportamentos mediante recompensas e castigos. Mas, além disto, sustentavam que para entender algo não se precisava estudar o cérebro, nem sentimentos tão complexos e subjetivos como a fome, o medo, ou o que elas significam. Desnecessário dizer que esta filosofia de pesquisa acabou se murchando durante a segunda metade do século XX, mas a inteligência artificial continuaria vigente por bem mais tempo.

Quando terminou a Segunda Guerra Mundial e se dispôs de computadores digitais para aplicações mais amplas, os pioneiros da inteligência artificial arregaçaram as mangas e começaram a programar. Tradução de línguas? Fácil! Não é mais do que decifrar um código. Só precisamos trocar cada símbolo do Sistema A por seu semelhante do Sistema B. Visão? Isso também parece fácil. Já conhecemos os teoremas geométricos que tratam de rotação, escala e deslocamento, e será simples codificá-los como algoritmos computacionais, de modo que nisto já temos meio caminho andado. Os experts em inteligência artificial realizaram grandes declarações sobre a rapidez com que a inteligência dos computadores primeiro igualaria, e depois, ultrapassaria a inteligência humana.

Parece irônico que o programa de computador que esteve mais próximo de superar o teste de Turing, um programa chamado Eliza, imitasse a um psicanalista, reformulando as perguntas que você fazia e as mandando de volta. Por exemplo, se uma pessoa digitasse: "Meu namorado e eu não estamos nos falando", Eliza diria: "Conta-me mais do seu namorado" ou "Por que pensa que seu namorado e você não se falam mais?". Projetado como uma brincadeira, o programa chegou a enganar a alguns, ainda que fosse tonto e superficial. Entre os esforços mais sérios incluíram-se programas como Blocks World, uma sala simulada que continha blocos de diferentes cores e formas. Você podia propor perguntas a Blocks World do tipo: "Há uma pirâmide verde sobre o cubo vermelho grande?", ou solicitar: "Põe o cubo azul sobre o cubo vermelho pequeno". O programa respondia a pergunta ou tratava de fazer o que você lhe pedia. Tudo era simulado, e

funcionava. Mas era limitado ao seu pequeno mundo artificial de blocos. Os programadores não foram capazes de generalizar para que fizesse algo útil.

Enquanto isso, o público vivia impressionado pela sucessão contínua de sucessos aparentes e novas histórias sobre a tecnologia da inteligência artificial. Um programa que gerou entusiasmo inicial era capaz de resolver teoremas matemáticos. Desde Platão, a inferência dedutiva de múltiplos passos era considerada o auge da inteligência humana, de modo que a princípio pareceu que a inteligência artificial tinha conseguido o grande prêmio. Mas, como no caso do Blocks World, acabou que o programa só podia achar teoremas muito simples que já eram conhecidos. Depois se suscitou uma grande agitação com os "sistemas experts", bases de dados que podiam responder perguntas propostas pelos usuários humanos. Por exemplo, um sistema expert médico seria capaz de diagnosticar a doença de um paciente se lhe dessem uma lista de sintomas. Mas novamente resultaram ser de uso limitado e não mostraram nada que se aproximasse de uma inteligência generalizada. Os computadores podiam jogar xadrez com perícia de experts, e o Deep Blue da IBM tornou-se famoso por acabar derrotando Gary Kasparov, o campeão mundial, em seu próprio jogo. Mas estes sucessos foram em vão. Deep Blue não ganhou por ser mais inteligente do que um ser humano, mas porque era milhões de vezes mais rápido. Deep Blue não tinha intuição. Um jogador expert humano olha uma posição do tabuleiro e de imediato vê que zonas de jogo têm maiores possibilidades de ser proveitosas ou perigosas, enquanto um computador não possui um sentido nato do que é importante e assim explora muito mais opções. Deep Blue também não tinha noção da história do jogo e não sabia nada do seu rival. Jogava xadrez, mas não o entendia, do mesmo modo que uma calculadora realiza operações aritméticas, mas não entende de matemática.

Em todos os casos, os programas de inteligência artificial que conseguiam sucesso só serviam para efetuar a tarefa particular especificada em seu projeto. Não generalizavam nem mostravam flexibilidade, e inclusive seus criadores admitiam que eles não pensavam como seres humanos. Alguns dos problemas da inteligência artificial que a princípio pareciam fáceis não conseguiram ser resolvidos. Na atualidade, nenhum computador pode compreender a linguagem tão bem como uma criança de três anos nem ver tão bem como um rato.

Depois de muitos anos de esforços, promessas não cumpridas e nenhum sucesso claro, a inteligência artificial começou a perder seu brilho. Os cientistas da disciplina passaram para outros campos de pesquisas. As empresas dedicadas a ela fracassaram e o financiamento se tornou mais escasso. Programar computadores para que realizassem as tarefas mais básicas de percepção, linguagem e comportamento começou a parecer impossível. Hoje a situação não mudou muito. Como já afirmei, ainda há pessoas que acham que os problemas da

inteligência artificial podem ser resolvidos com computadores mais rápidos, mas a maioria dos cientistas pensa que o esforço em si fracassou.

Não devemos culpar os pioneiros da inteligência artificial pelos seus fracassos. Alan Turing foi brilhante. Todos podiam dizer que a Máquina de Turing mudaria o mundo; e fez, mas não mediante a inteligência artificial.

* * *

Meu ceticismo para com as afirmações sobre a inteligência artificial aumentou na mesma época em que solicitei minha matrícula no MIT. John Searle, influente professor de filosofia da Universidade da Califórnia, em Berkeley, dedicava-se até então a sustentar que os computadores não eram e não podiam ser inteligentes. Para demonstrar isto, em 1980 ele apresentou um experimento mental chamado Sala Chinesa. Consistia no seguinte:

Suponhamos que haja uma sala com uma fenda em uma parede, e dentro da mesma se encontra um homem que fala inglês sentado diante de uma escrivaninha. Ele tem um grande livro de instruções e todo o lápis e papel de rascunho que ele possa precisar. Folheando o livro, ele vê que as instruções, escritas em inglês, ditam modos de manipular, classificar e comparar caracteres chineses. Mas, note, as instruções não dizem nada sobre o significado dos caracteres chineses; só tratam de instruir sobre como devem ser copiados, apagados, reordenados, transcritos e assim por diante.

Alguém do lado de fora da sala desliza um pedaço de papel pela fenda. Nele está escrito uma história e perguntas sobre a dita história, tudo em chinês. O homem lá dentro não fala nem lê uma única palavra de chinês, mas ele recolhe o papel e continua trabalhando com o livro de instruções. Às vezes estas lhe indicam para que ele escreva caracteres no papel de rascunho e outras para que ele mude e apague caracteres. Aplicando uma regra após outra, escrevendo e apagando caracteres, o homem se esforça até que o livro lhe anuncia que ele terminou. No fim das contas, ele escreveu uma nova página de caracteres que sem ele saber, são as respostas às perguntas. O livro indica-lhe para passar o papel pela fenda. Ele faz isto e se pergunta sobre a finalidade deste tedioso exercício.

Do lado de fora, uma pessoa que fala chinês lê a página. Ela nota que todas as respostas estão corretas; inclusive perspicazes. Se fosse perguntado a ela se essas respostas proviam de uma mente inteligente que tinha compreendido a história, ela diria: sem dúvida que sim. Mas está certa ela? Quem entendeu a história? Não foi a pessoa de dentro, certamente. Não foi o livro, que após tudo isto não é mais do que um volume colocado inerte sobre a escrivaninha entre montes de papel. Portanto, quando se efetuou o entendimento? A resposta de Searle é que não houve entendimento; só se folhearam muitas páginas sem sentido e algumas coisas foram copiadas e apagadas sem saber o que se fazia. E agora vem o

melhor: a sala chinesa é análoga a um computador digital. A pessoa é a CPU, que executa instruções sem pensar, e o papel de rascunho é a memória. Deste modo, por mais talento que se ponha no projeto de um computador para simular inteligência e produzir o mesmo comportamento de um ser humano, o dito computador não possuirá entendimento e não será inteligente. (Searle deixou claro que não sabia o que era a inteligência; ele só estava afirmando, que fosse o que fosse, os computadores não a tinham.)

Este raciocínio criou uma enorme disputa entre filósofos e experts em inteligência artificial. Gerou centenas de artigos, além de um pouco de sarcasmo e hostilidade. Os defensores da inteligência artificial contra-atacaram a Searle com dezenas de argumentos, declarando, entre outras coisas, que mesmo que nenhuma das partes componentes da sala entendesse chinês, a sala como um todo entendia, ou que a pessoa da sala entendia, ainda que não soubesse disto. Acho que Searle estava certo. Quando meditei sobre o raciocínio da sala chinesa e sobre o funcionamento dos computadores, não encontrei entendimento em nenhum lugar. Cheguei à convicção de que precisávamos compreender o que é o "entendimento", um modo de o definir de modo que pusesse de manifesto quando um sistema seria inteligente e quando não, quando se entenderia chinês e quando não. O comportamento não nos diz.

Um ser humano não precisa "fazer" nada para entender uma história. Posso ler uma história em silêncio, e ainda que eu não manifeste um comportamento evidente, meu entendimento e compreensão são claros, pelo menos para mim. Por outro lado, você não é capaz de afirmar por meu comportamento quieto se eu entendi ou não a história, nem sequer se conheço a língua em que está escrita. Você poderia propor-me depois perguntas para comprovar, mas meu entendimento ocorre quando leio a história, não quando respondo às suas perguntas. Este livro sustenta a tese de que o entendimento não pode ser medido pelo comportamento externo; como veremos nos próximos capítulos; é mais uma medição interna de como recorda as coisas o cérebro e de como ele emprega sua memória para fazer previsões. A sala chinesa, o *Deep Blue* e a maioria dos programas de computador não possuem nada semelhante. Não compreendem o que estão fazendo. O único modo de julgar se um computador é inteligente é por valorizar sua informação de saída, ou sua forma de trabalhar.

O argumento defensivo supremo da inteligência artificial é que os computadores, em teoria, poderiam simular o cérebro inteiro. Um computador poderia copiar o modelo de todos os neurônios e suas conexões, e neste caso não teria nada que distinguísse a "inteligência" do cérebro da "inteligência" de simulação por computador. Ainda que talvez pareça impossível de levar isto a cabo na prática, estou de acordo. Mas os pesquisadores da inteligência artificial não simulam cérebros, e seus programas não são inteligentes. Você não pode simular um cérebro sem primeiro compreender o que ele faz.

* * *

Depois de tanto a Intel como o MIT me recusarem, eu já não sabia mais o que fazer. E quando você não sabe como agir, a melhor estratégia costuma ser não efetuar mudanças até que se clareiem as opções. De modo que continuei trabalhando no campo da informática. Encontrava-me satisfeito em Boston, mas em 1982 minha esposa quis que nos mudássemos para a Califórnia, e assim o fizemos (era, mais uma vez, o caminho com menos obstáculos). Encontrei trabalho no Vale do Silício, em uma empresa de informática chamada Grid Systems. A dita empresa inventou o computador portátil, uma bela máquina que se converteu no primeiro computador da coleção do Museu de Arte Moderna de Nova York. Trabalhei primeiro no marketing e depois na engenharia, e acabei criando uma linguagem de programação de alto nível chamada GridTask. Ela e eu fomos nos tornando cada vez mais importantes para o sucesso da Grid; minha carreira ia bem.

Não obstante, não podia tirar da cabeça a curiosidade pelo cérebro e pelas máquinas inteligentes. Consumia-me o desejo de estudar os cérebros. De modo que fiz um curso por correspondência de psicologia humana e estudei por conta própria (ninguém jamais tinha sido recusado por uma escola de correspondência). Depois de aprender um bocado de biologia, decidi solicitar minha matrícula em um curso de pós-graduação de um programa de biologia e estudar a inteligência de dentro das ciências biológicas. Se a ciência da computação não queria um teórico do cérebro, talvez o mundo da biologia aceitasse um cientista dos computadores. Até então não existia biologia teórica nem nada semelhante, e muito menos neurociência teórica, assim biofísica parecia o melhor campo para meus interesses. Estudei muito, passei nos exames de seleção requeridos, preparei um currículo, pedi cartas de recomendação e consegui ser aceito como aluno de pós-graduação em tempo integral no programa de biofísica da Universidade da Califórnia, em Berkeley.

Estava emocionado. Finalmente eu poderia começar a sério com a teoria sobre o cérebro, ou pelo menos pensava. Deixei meu emprego na Grid sem intenção de voltar a trabalhar na indústria da informática. Certamente, isso significava renunciar indefinidamente a meu salário. Minha esposa achava que tinha chegado o momento de comprar uma casa e formar uma família, e eu me dispunha alegremente de deixar de contribuir com rendimentos. Este não era de modo algum o caminho com menos obstáculos, mas era a melhor opção que eu tinha, e ela apoiou minha decisão.

John Ellenby, fundador da Grid, enfiou-me em seu escritório justo antes que eu a deixasse e me disse: "Sei que você espera nunca mais voltar a Grid nem à indústria da informática, mas nunca se sabe o que pode acontecer. Em vez de deixar o emprego de vez, por que você não pede uma licença? Deste modo, se você regressar em um ano ou dois, você pode retomar seu salário, sua função e

sua opção sobre ações de onde os deixou". Foi um belo gesto, e aceitei; mas eu sentia que estava abandonando o mundo da informática para sempre.

Redes Neurais

Quando ingressei na Universidade de Berkeley, em janeiro de 1986, a primeira coisa que fiz foi compilar uma história das teorias sobre a inteligência e o funcionamento cerebral. Li centenas de artigos de anatomistas, fisiólogos, filósofos, linguistas, cientistas da computação e psicólogos. Numerosas pessoas de muitos campos tinham escrito extensamente sobre o pensamento e a inteligência. Cada campo contava com seu conjunto de publicações e cada um empregava sua própria terminologia. Suas descrições pareceram-me contraditórias e incompletas. Os linguistas falavam da inteligência com termos como "sintaxe" e "semântica". Para eles o cérebro e a inteligência só tratavam da linguagem. Os cientistas da visão faziam referência a esboços em 2D, 2 1/2 D e 3D. Para eles o cérebro e a inteligência consistiam no reconhecimento de padrões visuais. Os cientistas da computação falavam de esquemas e modelos, novos termos que tinham inventado para representar o conhecimento. Nenhuma destas pessoas ocupava-se de falar da estrutura do cérebro e de como isto levaria à prática qualquer uma das suas teorias. Por sua vez, os anatomistas e neurofisiólogos escreviam em abundância sobre a estrutura do cérebro e o comportamento dos neurônios, mas a maioria deles evitava qualquer tentativa de formular uma teoria em grande escala. Era difícil e frustrante tentar encontrar sentido nestas diversas propostas e na montanha de dados experimentais que as acompanhavam.

Nessa época apareceu em cena uma proposta inovadora e prometedora sobre as máquinas inteligentes. As redes neurais estavam presentes de uma forma ou de outra desde os finais da década de 1960, mas rivalizavam com o movimento da inteligência artificial pelo dinheiro e atenção dos organismos que financiavam as pesquisas. Durante vários anos, os pesquisadores das redes neurais estiveram em uma lista negra e não conseguiam fundos. No entanto, algumas pessoas continuaram se dedicando a elas, e em meados da década de 1980 chegou finalmente seu grande momento. É difícil saber com exatidão por que se teve um interesse repentino pelas redes neurais, mas sem dúvida um fator decisivo foi o fracasso contínuo da inteligência artificial. As pessoas buscavam alternativas e encontraram uma nelas.

As redes neurais representavam uma real melhora sobre a proposta da inteligência artificial porque sua arquitetura baseava-se, mesmo que em linhas muito gerais, em sistemas nervosos reais. Em vez de programar computadores, os pesquisadores das redes neurais, também conhecidos como *conexionistas*, se interessavam em aprender que tipos de comportamentos podiam se apresentar conectando juntos um monte de neurônios. Os cérebros estão compostos por neurônios; portanto, o cérebro é uma rede neural. Isso é um fato. Os conexionistas tinham a esperança de que as propriedades esquivas da inteligência se tornariam claras estudando como os neurônios interagem, e de que alguns dos

problemas insolúveis da inteligência artificial pudessem ser resolvidos reproduzindo as conexões precisas entre populações de neurônios. Uma rede neural se diferencia de um computador porque não tem CPU e não guarda informação em uma memória centralizada. O conhecimento e as memórias da rede distribuem-se por toda sua conectividade, assim como nos cérebros reais.

A primeira vista, as redes neurais pareciam se encaixar bem com meus interesses, mas esse campo logo acabou me desiludindo. Até então eu já tinha formado a opinião de que existia três coisas que pareciam essenciais para compreender o cérebro. Meu primeiro critério era a inclusão do tempo na função cerebral. Os cérebros reais processam com rapidez mudando fluxos de informação. Não há nada estático no fluxo de informação que entra e sai do cérebro.

O segundo critério era a importância da realimentação. Os neuroanatomistas sabiam desde há muito tempo que o cérebro está saturado de conexões de realimentação. Por exemplo, no circuito entre o neocórtex e uma estrutura inferior chamada tálamo, as conexões para atrás (de volta para a entrada) ultrapassam às que vão para adiante quase dez vezes, o que quer dizer que por cada fibra que nutre informação para adiante no neocórtex há dez fibras que nutrem informação para atrás aos sentidos. A realimentação domina também a maioria das conexões por todo o neocórtex. Ninguém compreende o papel preciso desta realimentação, mas pelas pesquisas publicadas parecia evidente que ela existia em todas as partes. Eu imaginava que ela devia ser importante.

O terceiro critério era que toda teoria ou modelo do cérebro deveria explicar a sua arquitetura física. O neocórtex não é uma estrutura simples. Como veremos mais adiante, está organizado em uma hierarquia que se repete. Sem dúvida, qualquer rede neural que não reconhecesse a dita estrutura não iria agir como um cérebro.

Mas quando o fenômeno das redes neurais ocupou a cena, ele se baseou em sua maioria em uma classe de modelos ultra simples que não cumpriam nenhum destes critérios. A maior parte das redes neurais constava de um pequeno número de neurônios ligados entre si em três filas. Apresentava-se um modelo (a entrada) à primeira fila. Estes neurônios de entrada estavam ligados com a próxima fila, as chamadas unidades ocultas. Depois estas se ligavam com a fila final de neurônios, as unidades de saída. As conexões entre os neurônios apresentavam forças variáveis, o que significava que a atividade em um neurônio poderia aumentar a atividade em outro e diminuir em um terceiro, dependendo das forças de conexão. Mudando as ditas forças, a rede aprenderia a representar modelos de entrada em modelos de saída.

Estas redes neurais simples só processavam modelos estáticos, não usavam realimentação e não se pareciam de forma alguma a cérebros. O tipo mais comum, chamado de rede de "propagação recorrente", aprendia transmitindo um erro das unidades de saída de volta para as unidades de entrada. Talvez você pense que

esta é uma forma de realimentação, mas não é. A propagação de erros para atrás só ocorria durante a fase de aprendizagem. Quando a rede neural funcionava com normalidade uma vez que tinha sido treinada; a informação só fluía em um sentido. Não tinha realimentação das saídas às entradas. E os modelos não tinham sentido do tempo. Um padrão de entrada estático transformava-se em um padrão de saída estático. Depois se apresentava outro padrão de entrada. Não se tinha histórico ou registro na rede do que tinha acontecido nem sequer um pouco antes. E, por último, a arquitetura destas redes neurais era trivial comparada com a estrutura complexa e hierárquica do cérebro.

Eu pensava que a disciplina passaria em seguida a se ocupar de redes mais realistas, mas não o fez. Já que estas redes neurais simples eram capazes de executar coisas interessantes, a pesquisa resolveu parar por aí durante anos. Tinham encontrado uma ferramenta nova e atraente, e da noite para o dia milhares de cientistas, engenheiros e estudantes obtinham bolsas, se doutoravam e escreviam livros sobre as redes neurais. Formaram-se empresas cujo fim era empregar as ditas redes para prognosticar os movimentos da Bolsa de Valores, processar solicitações de créditos, verificar assinaturas e realizar centenas de diversas aplicações de classificação de padrões. Ainda que a intenção dos fundadores do campo talvez fosse mais geral, ele acabou dominado por pessoas que não se interessavam pelo entendimento do funcionamento cerebral ou em que consistia a inteligência.

A imprensa popular não entendia muito bem a dita distinção. Os jornais, revistas e programas de ciência televisivos apresentavam as redes neurais como se fossem "semelhantes ao cérebro" ou funcionassem "segundo os mesmos princípios que o cérebro". Diferentes da inteligência artificial, onde tudo tinha que ser programado, as redes neurais aprendiam pelo exemplo, o que parecia de algum modo mais inteligente. Um exemplo destacado era a NetTalk. Esta rede neural aprendeu a representar sequências de letras sobre sons falados. Quando se exercitou a rede com um texto impresso, ela começou a soar como a voz de um computador lendo as palavras. Era fácil imaginar que com um pouco mais de tempo as redes neurais conversariam com os seres humanos. NetTalk foi anunciada erroneamente nas notícias nacionais como uma máquina que aprendia a ler. Era uma grande exibição, mas o que ela fazia na realidade beirava o trivial. Não lia, não entendia e não tinha nenhum valor prático. Limitava-se a casar combinações de letras com padrões de som predefinidos.

Permita-me citar uma analogia para mostrar o quão longe se achavam as redes neurais dos cérebros reais. Imaginemos que em vez de tentar pensar como funciona um cérebro estivéssemos tentando deduzir como funciona um computador digital. Depois de anos de estudo, descobrimos que tudo no computador está composto de transistores. Há centenas de milhões de transistores em um computador e os mesmos estão ligados entre si de formas precisas e

complexas. Mas não compreendemos como funciona o dito computador ou por que os transistores estão ligados desse modo. Portanto, em um dia decidimos ligar só alguns para ver o que acontecia. E, quem diria, descobrimos que quando interligamos apenas três transistores de uma determinada forma os mesmos se convertem em um amplificador. Um pequeno sinal colocado em um extremo é amplificado no outro. (Os amplificadores de rádios e TVs são realmente fabricados empregando transistores deste modo.) Trata-se de uma descoberta importante, e da noite para o dia surge uma indústria que se dedica a fabricar rádios, TVs e outros aparelhos de transistores utilizando os ditos amplificadores. Tudo isso é muito bom, mas não nos diz nada sobre como funciona o computador. Ainda que o amplificador e o computador sejam feitos de transistores, eles não têm quase mais nada em comum. Do mesmo modo, um cérebro real e a rede neural de três filas são constituídos de neurônios, mas não possuem quase mais nada em comum.

Durante o verão de 1987 tive uma experiência que jogou mais água fria sobre meu já escasso entusiasmo pelas redes neurais. Assisti a uma conferência sobre o tema em que vi uma apresentação de uma empresa chamada Nestor. A dita empresa tentava vender a aplicação de uma rede neural que reconhecia escrita manual sobre uma tabuleta. Ela oferecia a licença do programa por um milhão de dólares, o que me chamou a atenção. Ainda que a Nestor divulgasse a sofisticação do seu algoritmo de rede neural e o vendesse como outro importante avanço, me pareceu que o problema do reconhecimento de letras podia ser resolvido de uma forma mais simples e tradicional. Naquela noite voltei para casa pensando na questão, e em dois dias projetei um reconhecedor de letras mais rápido, mais pequeno e mais flexível. Não empregava uma rede neural e não funcionava de forma alguma como um cérebro, embora o algoritmo fosse inspirado em uma matemática que eu estava estudando relacionada aos cérebros. Essa conferência despertou meu interesse em projetar computadores com uma interface de ponteiro (o que acabou conduzindo à PalmPilot dez anos depois). O reconhecedor de letras que eu criei se converteu na base do sistema de entrada de texto chamado Graffiti, empregado na primeira série de produtos da Palm. Acredito que a Nestor tenha fechado as portas.

Era pedir muito para redes neurais tão simples. A maior parte das suas capacidades era suprida facilmente por outros métodos, e o entusiasmo com que as tinham recebido os meios de comunicação acabou diminuindo. Pelo menos, os pesquisadores das redes neurais não declaravam que seus modelos eram inteligentes. Além disso, tratava-se de redes extremamente simples e faziam menos coisas que os programas de inteligência artificial. Não quero lhe deixar a impressão de que todas as redes neurais pertencem à simples variedade de três camadas. Alguns pesquisadores continuam estudando redes com projetos diferentes. Na atualidade, o termo *rede neural* é empregado para descrever um conjunto diverso de modelos, alguns dos quais são mais precisos desde a

perspectiva biológica do que outros. Mas quase nenhum pretende captar a função ou a arquitetura geral do neocórtex.

Na minha opinião, o problema fundamental da maior parte das redes neurais é uma característica que é compartilhada com os programas de inteligência artificial. Ambas apoiam o ônus fatal de focar sua atenção no comportamento. Se eles estão chamando os ditos comportamentos de "respostas", "padrões" ou "saídas", tanto a inteligência artificial como as redes neurais presumem que a inteligência baseia-se no comportamento que um programa ou uma rede neural produz depois de processar uma determinada entrada. O atributo mais importante de um programa de computador ou uma rede neural é se os mesmos proporcionam a saída correta ou desejada. Como por inspiração de Alan Turing, inteligência é igual a comportamento.

Mas a inteligência não se reduz a agir ou a se comportar de modo inteligente. O comportamento é uma manifestação da inteligência, mas não a característica central ou a definição primordial de ser inteligente. Um momento de reflexão demonstra isto: você pode ser inteligente deitado na escuridão, pensando e compreendendo. Ignorar o que acontece *dentro* da sua cabeça e focar no comportamento tem constituído um grande impedimento para entender a inteligência e construir máquinas inteligentes.

* * *

Antes que exploremos uma nova definição da inteligência, quero lhe falar de outra proposta conexionista que esteve bem mais próxima de descrever como funcionam os cérebros reais. Pena que poucas pessoas parecem ter se dado conta da importância desta pesquisa.

Enquanto as redes neurais monopolizavam a atenção geral, um pequeno grupo desgarrado de teóricos do dito campo construía redes que não eram centradas no comportamento. Chamadas de memórias auto-associativas, as mesmas também eram formadas por "neurônios" simples que se ligavam entre si e se estimulavam quando atingiam certo limiar. Porém estavam interconectados de modo diferente, utilizando uma multidão de realimentações. Em vez de limitar-se a passar informação para adiante como em uma rede de propagação recorrente, as memórias auto-associativas alimentavam a saída de cada neurônio de volta para a entrada, algo parecido a chamar a si mesmo pelo telefone. Este emaranhado de realimentação conduziu a algumas características interessantes. Quando se imputava um padrão de atividade aos neurônios artificiais, os mesmos formavam uma memória do dito padrão. A rede auto-associativa associava padrões consigo mesma; daí o termo *memória auto-associativa*.

O resultado desta forma de conexão parece ridículo a princípio. Para recuperar um padrão armazenado na dita memória você deve proporcionar o padrão que você

quer recuperar. Seria como ir à quitanda para comprar uma dúzia de bananas. Quando o quitandeiro lhe perguntasse como você iria pagar, você responderia que pagaria com bananas. O que há de bom nisto?, você pode se perguntar. Porém a memória auto-associativa possui algumas propriedades importantes que se encontram nos cérebros reais.

A propriedade primordial é que você não precisa ter o padrão inteiro que você quer recuperar para poder fazer isto. Você poderia ter só parte do padrão ou um padrão desordenado. A memória auto-associativa pode recuperar o padrão correto tal como foi armazenado originalmente inclusive se você contribui com uma versão desordenada dele. Seria como ir a quitanda com umas bananas maduras e meio comidas e obter em troca bananas inteiras e verdes. Ou ir ao banco com uma cédula rasgada e ilegível e o caixa dissesse: "Acho que é uma cédula de 100 dólares danificada. Dê-me e lhe entregarei esta cédula nova de 100 dólares".

Em segundo lugar, diferente da maior parte das redes neurais restantes, pode-se projetar uma memória auto-associativa para que armazene sequências de padrões ou padrões temporais. Esta característica é conseguida acrescentando um *delay* temporal na realimentação. Com o dito *delay* você pode apresentar à uma memória auto-associativa uma sequência de padrões, similar a uma melodia, e ela será capaz de recordá-la. Eu poderia contribuir à primeira fila algumas notas de *Brilha, brilha, linda estrela*, e a memória devolveria a canção inteira. Quando se lhe apresenta parte da sequência, a memória é capaz de recordar o resto. Como veremos mais adiante, assim é como as pessoas aprendem quase tudo, como uma sequência de padrões. E proponho que o cérebro emprega circuitos similares a uma memória auto-associativa para fazer isto.

As memórias auto-associativas deram uma ideia da importância potencial que tinham as entradas com realimentação e mudança de tempo. Mas a grande maioria dos cientistas cognitivos, da inteligência artificial e das redes neurais passou por alto o tempo e a realimentação.

Em seu conjunto, os neurocientistas também não fizeram melhor. Também conheciam a realimentação, pois foram eles que a descobriram, mas a maioria carecia de teoria (além de uma vaga conversa sobre fases e modulação) para explicar por que o cérebro precisa tanto dela. E o tempo ocupa um papel escasso, quando ocupa, na maior parte das suas ideias sobre a função geral do cérebro. Tendem a representar o cérebro atendendo a onde ocorrem as coisas, não a quando e como os padrões neurais interagem ao longo do tempo. Parte disto provém das limitações as quais estão submetidas nossas técnicas experimentais atuais. Uma das tecnologias favoritas da década de 1990, também conhecida como a Década do Cérebro, foi a imagem funcional. As máquinas de imagem funcional podem tirar fotografias da atividade cerebral nos humanos, mas não são capazes de observar mudanças rápidas. Portanto, os cientistas solicitam aos sujeitos que se concentrem em uma única tarefa uma e outra vez como se fosse lhes pedido que

permanecessem quietos para uma fotografia óptica, com a advertência de que a mesma seria uma fotografia mental. Como resultado, contamos com uma multidão de dados sobre onde ocorrem certas tarefas no cérebro, mas poucos sobre como fluem por ele entradas reais que variam com o tempo. A imagem funcional oferece a oportunidade de ver aonde acontecem as coisas em um determinado momento, mas não é capaz de captar facilmente como a atividade cerebral muda ao longo do tempo. Os cientistas desejariam reunir estes dados, mas existem poucas boas técnicas para conseguir isto. Deste modo, muitos neurocientistas cognitivos da corrente dominante continuam participando na falácia entrada-saída. Você apresenta uma determinada entrada e vê que saída você obtém. Os diagramas de conexão do córtex tendem a mostrar mapas de fluxos que começam nas áreas sensoriais primárias onde entram a visão, os sons e o tato, fluem por áreas analíticas, planejadoras e motoras superiores, e depois passam a alimentar instruções aos músculos. Você sente, e depois age.

Não quero dar a entender que ninguém tenha tido em conta o tempo e a realimentação. Trata-se de um campo tão enorme que quase qualquer ideia conta com seus partidários. Nos anos recentes houve um aumento na crença da importância da realimentação, do tempo e da predição. Mas o estrondo da inteligência artificial e das redes neurais clássicas manteve subjugadas e depreciadas as propostas restantes durante anos.

* * *

Não é difícil entender por que as pessoas —tanto leigos como experts— pensam que o comportamento define a inteligência. Durante um par de séculos ou menos, as capacidades do cérebro foram comparadas com os mecanismos do relógio, a seguir com bombas e tubulações, depois com motores à vapor e mais tarde com computadores. Décadas de ficção científica têm transbordado de ideias sobre a inteligência artificial, das leis da robótica de Isaac Asimov ao C3PO de *Guerra das estrelas*. A ideia de máquinas que *fazem* coisas está arraigada em nossa imaginação. Todas as máquinas, sejam fabricadas ou imaginadas pelos humanos, são projetadas para fazer algo. Não temos máquinas que pensam; temos máquinas que fazem. Inclusive quando observamos a nossos semelhantes humanos, nos centramos em seu comportamento e não em seus pensamentos ocultos. Portanto, parece intuitivamente óbvio que o comportamento inteligente deve ser a medida de um sistema inteligente.

No entanto, quando se observa a história da ciência, comprova-se que nossa intuição costuma ser o maior obstáculo para descobrir a verdade. Os modelos científicos são com frequência difíceis de serem descobertos, não porque sejam complexos, mas porque as hipóteses intuitivas, porém errôneas, nos impedem ver a resposta correta. Os astrônomos anteriores a Copérnico (1473-1543) erraram ao supor que a Terra permanecia quieta no centro do Universo porque *parecia* estar quieta e o ocupar. Era uma intuição evidente que todas as estrelas faziam parte de

uma esfera gigantesca com a gente no centro. Sugerir que a Terra girava como um pião, com sua superfície se movendo a mais de 1.600 quilômetros por hora, e que ela se deslocava rapidamente pelo espaço —sem mencionar que as estrelas se encontravam a trilhões de quilômetros de distância—, levaria você a ser considerado como um lunático. Mas acabou que esse era o modelo correto. Simples de entender, mas errôneo segundo a intuição.

Antes de Darwin parecia evidente que as espécies possuíam formas fixas. Os crocodilos não se parecem com os colibris; são diferentes e irreconciliáveis. A ideia de que as espécies evoluem não só ia contra os ensinamentos religiosos, mas também contra o senso comum. A evolução afirma que temos um antepassado comum com qualquer ser vivo deste planeta, incluindo os vermes e a planta florida da nossa cozinha. Agora sabemos o que *talvez* seja verdade, porém a intuição diz o contrário.

Menciono estes exemplos famosos porque acho que a busca de máquinas inteligentes também apoia o ônus de uma hipótese intuitiva que está dificultando nosso progresso. Quando você se pergunta o que faz um sistema inteligente, é evidente que a intuição dita pensar no comportamento. Demonstramos a inteligência humana mediante a fala, a escrita e as ações, não é verdade? Sim; mas só até certo ponto. A inteligência é algo que acontece em nossa cabeça. O comportamento é um ingrediente opcional. Isto não parece óbvio segundo a intuição, mas também não é difícil de entender.

* * *

Na primavera de 1986, enquanto sentava-me diante da minha escrivaninha em um dia após outro lendo artigos científicos, construindo minha história da inteligência e observando os mundos em evolução da inteligência artificial e das redes neurais, me encontrei afogado em detalhes. Eu tinha um suprimento interminável de coisas a ler e estudar, mas não estava conseguindo nenhum entendimento claro de como funcionava na realidade o cérebro como um todo, nem sequer do que ele fazia. Isso se devia por que o próprio campo da neurociência estava inundado de detalhes. E continua estando. A cada ano publicam-se centenas de relatórios de pesquisas, porém tendem a acrescentar ao montão em vez de organizá-lo. Ainda não existe uma teoria geral, um modelo, que explique o que faz nosso cérebro e por que.

Comecei a imaginar como seria a solução para este problema. Seria extremamente complexa porque o cérebro é muito complexo? Precitaria de cem páginas de matemática densa para descrever como funciona o cérebro? Seria necessário representar centenas ou milhares de circuitos separados antes que se pudesse compreender algo útil? Eu achava que não. A história mostra que as melhores soluções aos problemas científicos são simples e elegantes. Ainda que os detalhes

pareçam intimidantes e o caminho para a teoria final seja árduo, por regra geral o modelo conceitual definitivo é simples.

Sem uma explicação central que guie a indagação, os neurocientistas não contam com muito para prosseguir enquanto tentam ajuntar todos os detalhes que têm reunido para formar um quadro coerente. O cérebro é incrivelmente complexo, um emaranhado de células vasto e surpreendente. A primeira vista parece um estádio cheio de espaguete cozido. Também já foi descrito como o pesadelo de um eletricitista. Mas depois de uma inspeção minuciosa vemos que o cérebro não é um montão aleatório. Possui muita organização e estrutura, porém o suficiente para que possamos esperar ser capazes de chegar a intuir o funcionamento do conjunto da mesma forma que somos capazes de ver como os fragmentos de um vaso quebrado voltam a se unir. A falha não consiste em carecer de dados suficientes ou de dados precisos; o que precisamos é de uma mudança de perspectiva. Com o modelo adequado os detalhes ganharão significado e se tornarão manejáveis. Consideremos a seguinte analogia imaginária para você conseguir apreciar o que quero dizer.

Imaginemos que daqui à vários milênios os humanos se extingam e que exploradores de uma civilização extraterrestre avançada desembarquem na Terra. Eles querem deduzir como vivíamos. Intriga-lhes em particular nossas redes de estradas. Para que serviam essas estruturas elaboradas e estranhas? Começam catalogando todas elas, tanto via satélite como desde o solo. São arqueólogos meticolosos. Registram a situação de cada fragmento perdido de asfalto, cada placa de sinalização caída e arrastada morro abaixo pela erosão, cada detalhe que possam encontrar. Dão-se conta de que algumas redes de estradas são diferentes de outras; em certos lugares são sinuosas e estreitas, e sua aparência é quase aleatória; em outros formam uma rede regular e em alguns trechos se tornam densas e percorrem centenas de quilômetros pelo deserto. Eles recolhem uma montanha de detalhes, mas que não significam nada para eles. Continuam reunindo mais e mais com a esperança de encontrar algum dado novo que explique tudo. Continuam perplexos durante muito tempo.

E assim permanecem, até que um deles exclama: "Eureca! Acho que descobri... essas criaturas não podiam teletransportar-se como nós. Tinham que viajar de um lugar a outro, talvez sobre plataformas móveis com um projeto engenhoso". A partir desta percepção básica, muitos detalhes começam a ficar claros. As redes de ruas pequenas e sinuosas correspondem às primeiras épocas, quando os meios de transporte eram lentos. As autopistas densas e longas serviam para percorrer grandes distâncias à velocidades elevadas e as mesmas sugeriam finalmente uma explicação do porquê dos sinais dessas estradas terem números diferentes pintados. Os cientistas começam a deduzir as zonas residenciais das industriais, a forma em que as demandas do comércio e a infraestrutura de transporte deviam ter interagido, e assim sucessivamente. Muitos dos detalhes que eles tinham

catalogado acabaram não sendo muito importantes; apenas acidentes da história ou características da geografia local. Existe a mesma quantidade de dados brutos, mas já não são desconcertantes.

Cabe confiar que o mesmo tipo de avanço nos permitirá compreender o que significam todos os detalhes do cérebro.

* * *

Infelizmente, nem todos acham que sejamos capazes de entender como funciona o cérebro. Um surpreendente número de pessoas, incluindo alguns neurocientistas, acha que de certo modo o cérebro e a inteligência estão além de qualquer explicação. E alguns acham que, ainda que conseguíssemos os compreender, seria impossível construir máquinas que funcionem do mesmo modo, por que a inteligência requer um corpo humano, neurônios, e talvez algumas novas e insondáveis leis da física. Sempre que escuto estes argumentos, imagino os intelectuais do passado que se negavam ao estudo do céu ou se opunham à dissecação dos cadáveres para ver como funcionavam nossos corpos. “Não se preocupe em estudar isso; não levará a nada de bom e, ainda que você conseguisse compreender como funciona, não há nada que se possa fazer com esse conhecimento.” Raciocínios como este nos conduzem a um ramo da filosofia chamado funcionalismo, nossa última parada nesta breve história da reflexão sobre o pensamento.

Segundo o funcionalismo, ser inteligente ou ter uma mente não é mais do que uma propriedade organizativa, e em essência carece de importância do que você seja composto. Existe uma mente em todo sistema cujas partes constituintes possuem a relação causal mútua adequada, mas essas partes podem ser com a mesma validade neurônios, chips de silício ou outra coisa. Sem dúvida, esta opinião é normal para qualquer aspirante à construtor de máquinas inteligentes.

Consideremos o seguinte: Um jogo de xadrez seria menos real se fosse jogado com um saleiro substituindo a peça perdida de um cavalo? É evidente que não. O saleiro é o equivalente funcional de um cavalo “real” em virtude de como ele se move sobre o tabuleiro e interage com as peças restantes, de modo que se trata de um jogo de xadrez de verdade e não de uma simulação. Ou consideremos se esta oração seria a mesma se eu apagasse com meu cursor cada um dos caracteres e voltasse a digitar. Ou, tomemos um exemplo mais familiar, consideremos o fato de que todos os anos nosso corpo substitui a maioria dos átomos que nos compõem. Apesar disso, continuamos sendo os mesmos em todos os sentidos que nos importam. Um átomo é tão bom como qualquer outro se ele desempenha o mesmo papel funcional em nossa constituição molecular. Cabe sustentar o mesmo no caso do cérebro: se um cientista louco substituísse cada um dos nossos neurônios com uma réplica nana-robótica funcionalmente equivalente,

deveríamos sair do processo sentindo-nos não menos que nós mesmos do que nos sentíamos no começo.

Por este princípio, um sistema artificial que empregue a mesma arquitetura funcional de um cérebro vivo inteligente deve ser também inteligente e não se limitar a aparentar; deve ser real, realmente inteligente.

Os defensores da inteligência artificial, os conexionistas e eu somos funcionalistas na medida em que todos nós achamos que não há nada inerente, especial ou mágico no cérebro que lhe permita ser inteligente. Todos nós achamos que seremos capazes de construir máquinas inteligentes de certo modo algum dia. Mas existem interpretações diferentes do funcionalismo. Ainda que eu já tenha declarado qual considero ser a falha central da inteligência artificial e dos paradigmas conexionistas —a falácia entrada-saída—, vale a pena dizer algo mais a respeito do por que ainda não temos sido capazes de projetar máquinas inteligentes. Enquanto os defensores da inteligência artificial adotam o que considero ser uma linha dura autodestrutiva, na minha opinião os conexionistas têm sido simplesmente muito tímidos.

Os pesquisadores da inteligência artificial perguntam: “Por que nós engenheiros devemos estar limitados pelas soluções com as quais a evolução deu por acaso?”. Em princípio, eles têm razão. Os sistemas biológicos, como o cérebro e o genoma, são considerados muito pouco elegantes. Uma metáfora habitual é a da máquina de Rube Goldberg, batizada com o nome do caricaturista da Era da Depressão que desenhava artefatos cômicos extremamente complexos para realizar tarefas triviais. Os projetistas de software contam com um termo relacionado, *kludge*, para referir-se a programas escritos sem planejamento que acabam sendo repletos de uma complexidade onerosa e inútil, frequentemente chegando até o ponto de se tornar incompreensíveis inclusive para os programadores que os escreveram. Os pesquisadores da inteligência artificial temem que o cérebro seja uma confusão similar, um *kludge* de vários milhões de anos cheio até o topo de ineficiências e um “código herdado” evolucionista. Se for assim, eles se perguntam, por que não nos desfazer de toda essa penosa bagunça e começar novamente?

Muitos filósofos e psicólogos cognitivos mostram-se favoráveis a esta postura. Gostam da metáfora de que a mente se assemelha a um *software* que põe em funcionamento o cérebro, o análogo orgânico do *hardware* de computador. Nos computadores, os níveis de hardware e *software* são diferentes um do outro. O mesmo programa de software funciona em qualquer Máquina de Turing Universal. Você pode utilizar WordPerfect em um computador pessoal, em um Macintosh ou em um supercomputador Cray, por exemplo, ainda que estes três sistemas possuam diferentes configurações de hardware. E o *hardware* não tem nada importante a nos ensinar se estamos tentando aprender WordPerfect. Por analogia, prossegue o raciocínio, o cérebro não tem nada a ensinar sobre a mente.

Os defensores da inteligência artificial também gostam de assinalar exemplos históricos no qual a solução da engenharia difere radicalmente da versão da Natureza. Por exemplo, como conseguimos construir máquinas voadoras?, imitando o movimento de bater asas dos animais alados? Não. Fizemos com asas fixas e hélices, e mais adiante com motores propulsores. Pode ser que não seja como fez a Natureza, mas funciona, e melhor do que batendo asas.

De modo similar, fizemos um veículo de terra que podia correr mais do que os guepardos, não fabricando máquinas de quatro patas semelhantes aos ditos animais, mas inventando as rodas. É um modo excelente de se mover sobre o terreno plano, e o fato de que a evolução nunca tenha se topado com esta estratégia particular não significa que ela não seja uma ótima forma de nos deslocar. Alguns filósofos da mente ficaram vibrados com a metáfora da "roda cognitiva", isto é, a solução da inteligência artificial para algum problema que, ainda que seja completamente diferente de como o faz o cérebro, o faz bem. Em outras palavras, um programa que produz saídas que se parecem (ou superam) a execução humana de uma tarefa de modo limitado, porém útil, é tão bom quanto a forma em que nossos cérebros o fazem.

Acho que este tipo de interpretação de fins que justificam os meios do funcionalismo desencaminhou os pesquisadores da inteligência artificial. Como demonstrou Searle com a sala chinesa, não basta a equivalência funcional. Já que a inteligência é uma propriedade interna do cérebro, temos que olhar dentro dele para entender o que ela é. Em nossas pesquisas sobre o cérebro, e em especial do neocórtex, precisaremos ser minuciosos ao elucidar quais detalhes são apenas "acidentes congelados" supérfluos do nosso passado evolutivo; sem dúvida, muitos processos do tipo Rube Goldberg estão misturados com as características importantes. Mas, como veremos em seguida, existe uma elegância subjacente de grande potência, uma que supera nossos melhores computadores, esperando ser extraída desses circuitos neurais.

Os conexionistas perceberam intuitivamente que o cérebro não era um computador e que seus segredos se arraigavam no modo de se comportar seus neurônios quando se ligavam entre si. Foi um bom começo, mas o campo mal tem avançado desde seus primeiros lucros. Ainda que milhares de pessoas tenham trabalhado em redes de três camadas, e muitas ainda continuam fazendo isto, as pesquisas sobre redes corticais realistas foram e continuam sendo raras.

Durante meio século estivemos aplicando toda a força do considerável talento da nossa espécie para tentar programar inteligência nos computadores. No processo nos veio processadores de texto, bases de dados, videogames, Internet, telefones móveis e convincentes dinossauros animados por computador. Mas as máquinas inteligentes continuam não aparecendo no quadro. Para se ter sucesso precisaremos copiar muito do motor da inteligência da Natureza, o neocórtex.

Temos que extrair a inteligência do cérebro. Nenhum outro caminho nos conduzirá até elas.

3

O Cérebro Humano

Portanto, o que torna o cérebro tão diferente da programação que se incorpora na inteligência artificial e nas redes neurais? O que tem de inusual o projeto do cérebro e por que é importante? Como veremos nos próximos capítulos, a arquitetura cerebral tem muito o que nos dizer sobre como funciona o cérebro e por que ele é radicalmente diferente de um computador.

Começamos nossa introdução com o órgão em seu conjunto. Imaginemos que há um cérebro colocado sobre uma mesa e que o dissecamos juntos. A primeira coisa que apreciamos é que sua superfície exterior parece muito uniforme. De um cinza-rosado, ela assemelha-se a uma suave couve-flor, com numerosas cristas e vales, chamados de circunvoluções e sulcos. É macia e úmida ao tato. Esta superfície trata-se do neocórtex, uma fina camada de tecido neural que envolve a maioria das partes mais antigas do cérebro. Nele vamos focar nossa atenção particular. Quase tudo o que pensamos que é inteligência —a percepção, a linguagem, a imaginação, a matemática, a arte, a música e o planejamento— ocorre nele. Seu neocórtex está lendo este livro.

Agora tenho que admitir que sou um chauvinista neocortical. Sei que vou encontrar certa resistência por isso, de modo que me permita um minuto para defender minha postura antes de prosseguir a explicando. Cada parte do cérebro possui sua própria comunidade de cientistas que a estudam, e sugerir que possamos chegar na base da inteligência entendendo somente o neocórtex sem dúvida suscitará alguns alaridos e objeções das comunidades de pesquisadores ofendidos. Dirão coisas como: “Não é possível entender o neocórtex sem compreender a região cerebral *tal*, porque as duas estão muito interconectadas, e precisa-se da região cerebral *tal* para fazer isto e aquilo outro”. Estou de acordo. Concordo que o cérebro consta de muitas partes e a maioria é crucial para o ser humano. (Uma curiosa exceção é a parte do cérebro com o maior número de células, o cerebelo. Se você nasce sem cerebelo ou defeituoso, você pode levar uma vida normal. No entanto, não ocorre o mesmo com a maioria das regiões restantes, que se requerem para a vida básica ou para o estado consciente.)

Meu raciocínio é que não estou interessado em construir humanos. Quero entender a inteligência e construir máquinas inteligentes. Ser humano e ser inteligente são assuntos separados. Uma máquina inteligente não precisa ter impulsos sexuais, fome, pulsação, músculos, emoções ou corpo semelhante ao humano. Um humano é bem mais do que uma máquina inteligente. Somos criaturas biológicas com todo o necessário e às vezes com bagagem indesejada proveniente de eras de evolução. Se você deseja construir máquinas inteligentes que se comportem como humanos —isto é, que passem no teste de Turing em todos seus aspectos— é provável que você tenha que recriar boa parte da restante composição que torna

os humanos o que são. Mas, como veremos mais adiante, para construir máquinas que sejam inteligentes para valer, mas não exatamente iguais aos humanos, podemos focar na parte do cérebro estritamente relacionada com a inteligência.

Àqueles que talvez se sintam ofendidos pela atenção singular que presto ao neocórtex, lhes direi que estou de acordo que outras estruturas cerebrais, como o tronco encefálico, os gânglios basais e o núcleo amigdalino, são importantes para o funcionamento do mesmo. Sem dúvida alguma. Mas espero convencê-lo de que todos os aspectos essenciais da inteligência ocorrem no neocórtex, ainda que outras regiões cerebrais também possam desempenhar importantes papéis, como o tálamo e o hipocampo, dos quais nos ocuparemos mais adiante no livro. A longo prazo, precisaremos compreender os papéis funcionais de todas as regiões cerebrais. Mas acho que esses temas se abordarão melhor no contexto de uma boa teoria geral da função neocortical. Esta é minha opinião resumida. Agora voltemos ao neocórtex, ou para encurtar, o córtex.

Apanhe seis cartões de visita ou cartas de baralho —quaisquer um deles valerá— e os ponha em um monte. (Conviria que você fizesse isto para valer em vez de se limitar a imaginar.) Agora você conta com um modelo do córtex. Os seis cartões têm uma espessura de uns dois milímetros e lhe proporcionará o sentido de quão fina que é a lâmina cortical. Assim como o monte de cartões ou cartas, o neocórtex tem uma espessura aproximada de dois milímetros e conta com seis camadas, mais ou menos uma por cada cartão ou carta.

Estendida, a lâmina neocortical humana atinge o tamanho aproximado de um guardanapo grande. As lâminas corticais de outros mamíferos são menores: a do rato é do tamanho de um selo de correios; a do macaco, do tamanho de um envelope de uma carta comercial. Mas, deixando de lado o tamanho, a maioria delas contém seis camadas similares às que vemos no monte de cartões de visita. Os humanos são mais espertos porque nosso córtex, em relação ao tamanho corporal, ocupa uma zona maior, e não porque nossas camadas sejam mais grossas ou contenham alguma classe especial de células “espertas”. Seu tamanho é bastante impressionante, pois rodeia e envolve a maior parte do resto do cérebro. Para acomodar nosso grande cérebro, a Natureza teve que modificar nossa anatomia geral. As fêmeas humanas desenvolveram uma pélvis larga para dar a luz crianças de cabeça grande, característica que alguns paleoantropólogos pensam que coevoluiu com a capacidade de caminhar sobre as duas pernas. Mas não sendo suficiente, a evolução dobrou o neocórtex, o enfiando dentro dos nossos crânios como uma folha de papel enrugada dentro de um copo de conhaque.

O neocórtex está carregado de células nervosas ou neurônios. Estão tão abarrotadas que ninguém sabe com precisão quantas células ele contém. Se você desenhar um quadrado diminuto de um milímetro de lado (aproximadamente a metade do tamanho desta letra *o*) na parte superior do monte de cartões de visita,

você estará marcando a posição estimada de cem mil neurônios. Imaginem tentar contar o número exato em um espaço tão reduzido; é quase impossível. Não obstante, alguns anatomistas têm calculado que o neocórtex humano médio contém cerca de trinta bilhões de neurônios, mas ninguém se surpreenderia se a cifra fosse bem mais alta ou bem baixa.

Esses trinta bilhões de células é você. Eles contêm quase todas as suas lembranças, conhecimentos, capacidades e experiência vital acumulada. Depois de vinte e cinco anos pensando sobre cérebros, este fato continua-me parecendo espantoso. Que uma fina lâmina de células veja, sinta e crie nossa visão do mundo é algo incrível. O calor de um dia de verão e os sonhos que temos de um mundo melhor são de certo modo a criação destas células. Muitos anos após ele ter publicado seu artigo em *Scientific American*, Francis Crick escreveu um livro sobre cérebros chamado *The Astonishing Hypothesis*. A hipótese espantosa era simplesmente que a mente é a criação das células do cérebro. Não há nada mais, nada mágico, nenhum molho especial; apenas neurônios e uma dança de informações. Espero que você seja capaz de perceber como é incrível se dar conta disso. Parece existir um grande abismo filosófico entre uma reunião de células e nossa experiência consciente, embora a mente e o cérebro sejam a mesma coisa. Ao chamá-la de uma hipótese, Crick mostrava-se politicamente correto. Que as células do nosso cérebro criam a mente é um fato, não uma hipótese. Precisamos compreender o que fazem esses trinta bilhões de células e como o fazem. Por sorte, o córtex não é só uma bolha amorfa de células. Podemos observá-lo em profundidade para buscar ideias sobre como se comporta a mente humana.

* * *

Voltemos a nossa mesa de dissecação e olhemos um pouco mais o cérebro. A olho nu, o neocórtex quase não oferece sinais. Sem dúvida, há alguns, como a enorme fissura que separa os dois hemisférios cerebrais e o sulco proeminente que divide as regiões posteriores e frontais. Mas onde quer que você olhe, da esquerda para a direita e de trás para a frente, a superfície enrolada parece muito parecida. Não existem linhas limítrofes visíveis ou códigos de cor que delimitem zonas especializadas em diferentes informações sensoriais ou diferentes tipos de pensamento.

No entanto, nós sabemos desde há muito tempo que existe algum tipo de limite. Inclusive antes que os neurocientistas fossem capazes de discernir algo útil sobre o sistema de circuitos do córtex, eles sabiam que algumas funções mentais estavam localizadas em certas regiões. Se uma apoplexia deixa fora de combate o lóbulo parietal direito de Joe, ele pode perder sua capacidade de perceber —ou inclusive de conceber— qualquer coisa do lado esquerdo do seu corpo, ou da metade esquerda do espaço ao seu redor. Em contraste, uma apoplexia na região frontal esquerda, conhecida como área de Broca, compromete sua capacidade de empregar as regras gramaticais, ainda que seu vocabulário e sua faculdade para

entender os significados das palavras não mudem. Uma apoplexia em uma zona chamada circunvolução fusiforme pode acabar com a capacidade de reconhecer rostos: Joe pode não reconhecer a sua mãe, seus filhos e nem sequer seu próprio rosto em uma fotografia. Transtornos tão fascinantes como estes facilitaram cedo aos neurocientistas a noção de que o córtex consta de muitas regiões ou áreas funcionais. Os termos são equivalentes.

Aprendemos bastante sobre áreas funcionais no século passado, mas ainda há muito por descobrir. Cada uma destas regiões é semi-independente e parece ser especializada em certos aspectos da percepção ou do pensamento. Fisicamente, estão dispostas como uma série de remendos irregulares que varia um pouco de uma pessoa a outra. Raramente as funções estão delimitadas com clareza. Desde a perspectiva funcional, estão organizadas em uma hierarquia com ramificações.

A noção de hierarquia é crucial, de modo que quero dedicar certo tempo para defini-la com cuidado, pois me estarei referindo a ela ao longo de todo o livro. Em um sistema hierárquico, alguns elementos estão em um sentido abstrato “acima” e “abaixo” dos outros. Em uma hierarquia empresarial, por exemplo, uma gerente de nível médio está acima do empregado que se encarrega do correio e abaixo do vice-presidente. Isto não tem nada que ver com estar acima ou abaixo fisicamente; ainda que ela trabalhe em um andar inferior ao do encarregado do correio, a gerente continua estando “acima” desde o ponto de vista hierárquico. Enfatizo este ponto para tornar claro o que quero dizer quando falo que uma região funcional está mais alta ou mais baixa que outra. Não tem nada a ver com sua disposição física no cérebro. Todas as áreas funcionais do córtex residem na mesma lâmina cortical. O que faz com que uma região seja “superior” ou “inferior” a outra é a forma delas se ligarem entre si. No córtex, as áreas inferiores alimentam informação às superiores mediante um padrão neural de conectividade, enquanto as áreas superiores enviam realimentação “para abaixo” às áreas inferiores empregando um padrão de conexão diferente. Também existem conexões laterais entre áreas que estão em ramos separados da hierarquia, do mesmo modo que um gerente de nível médio se comunica com seu semelhante em um escritório associado de outro estado. Dois cientistas, Daniel Felleman e David von Essen, elaboraram um mapa detalhado do córtex do macaco. O dito mapa mostra dezenas de regiões ligadas entre si em uma hierarquia complexa. Cabe assumir que o córtex humano possui uma hierarquia similar.

As regiões funcionais mais baixas, as áreas sensoriais primárias, é o lugar aonde chega primeiro a informação sensorial. Elas processam a informação em seu nível mais bruto e básico. Por exemplo, a informação visual entra no córtex através da área visual primária, chamada de V1 para abreviar. V1 está ligada com características visuais de nível inferior, como segmentos de bordas diminutas, componentes de pequena escala do movimento, disparidade binocular (para a visão estérea) e informação básica de cor e contraste. V1 fornece informação às

áreas V2, V4 e IT (falaremos sobre elas mais adiante), e à muitas outras regiões. Cada uma das ditas áreas ocupa-se de aspectos mais especializados ou abstratos da informação. Por exemplo, as células de V4 correspondem a objetos de complexidade média, como formas de estrela em diferentes cores como vermelho ou azul. Outra área chamada MT é especializada nos movimentos dos objetos. Nos degraus mais elevados do córtex visual encontram-se áreas que representam lembranças visuais de todo tipo de objetos, como rostos, animais, ferramentas, partes do corpo e assim por diante.

Os sentidos restantes apresentam hierarquias similares. O córtex possui uma área auditiva primária chamada A1 e uma hierarquia de regiões auditivas acima, e conta com uma área somatosensorial primária (sentido corporal) chamada S1 e uma hierarquia de regiões somatosensoriais acima. Por fim, a informação sensorial passa para as "áreas de associação", que é o nome que às vezes se emprega para as regiões do córtex que recebem entradas de mais de um sentido. Por exemplo, nosso córtex tem áreas que recebem entradas tanto da visão como do tato. Graças às regiões de associação somos capazes de dar-nos conta de que a visão de uma mosca andando por nosso braço e a sensação de cócegas que sentimos compartilham a mesma causa. A maioria destas áreas recebe entradas altamente processadas de vários sentidos, e suas funções continuam sem estar claras. Mais adiante no livro terei muito o que dizer sobre a hierarquia cortical.

Existe mais um conjunto de áreas nos lóbulos frontais do cérebro que criam saídas motoras. O sistema motor do córtex também está organizado segundo uma hierarquia. A área inferior, M1, envia conexões à medula espinhal e maneja os músculos de forma direta. As áreas superiores alimentam de ordens motoras complexas à M1. A hierarquia da área motora e as hierarquias das áreas sensoriais parecem ser muito similares. Parecem estar organizadas da mesma forma. Na região motora pensamos na informação que flui para abaixo da hierarquia até M1 para manejar os músculos, e nas regiões sensoriais pensamos na informação que flui para acima da hierarquia se afastando dos sentidos. Mas na realidade a informação flui em ambas as direções. O que nas regiões sensoriais se entende como realimentação na região motora se entende como saída, e vice-versa.

A maioria das descrições dos cérebros é baseada em mapas de fluxos que refletem uma visão das hierarquias muito simplificada. Isto é, a entrada (visões, sons, tatos) flui às áreas sensoriais primárias e é processada enquanto avança para acima da hierarquia, depois passa pelas áreas de associação, a seguir pelos lóbulos frontais do córtex, e por último desce às áreas motoras. Não afirmo que esta visão seja completamente equivocada. Quando lemos em voz alta, a informação visual entra em V1, flui até as áreas de associação, faz sua rota até o córtex motor frontal e termina fazendo com que os músculos da nossa boca e garganta formem os sons da fala. No entanto, isto não é tudo o que ocorre. Não é tão simples. Na visão muito simplificada contra a qual previno, o processo se costuma tratar como

se a informação fluísse em uma única direção, como as peças que se armam na corrente de montagem de uma fábrica. Mas a informação no córtex sempre flui também na direção contrária, e com muito mais projeções alimentando para abaixo da hierarquia do que para acima. Quando lemos em voz alta, as regiões superiores do nosso córtex enviam mais sinais para “abaixo” do nosso córtex visual primário do que os que nosso olho recebe da página impressa. Nos próximos capítulos iremos nos ocupar do que essas projeções de realimentação fazem. Por enquanto, quero que você grave um fato: ainda que a hierarquia ascendente seja real, temos que ter cuidado para não pensar que o fluxo de informação só tem uma direção.

Novamente na mesa de dissecação, suponhamos que instalamos um potente microscópio, cortamos uma fatia fina da lâmina cortical, tingimos algumas células e damos uma olhada na nossa obra através da ocular. Se tingimos todas as células da nossa fatia, veremos uma massa toda negra porque as células estão muito abarrotadas e entrelaçadas. Mas se empregamos uma tinta que marque uma fração menor de células, podemos ver as seis camadas que tenho mencionado. Estas camadas estão formadas por variações da densidade das células corporais, dos tipos de células e das suas conexões.

Todos os neurônios possuem características em comum. Além do corpo celular, que é a parte arredondada que imaginamos quando pensamos em uma célula, elas também possuem estruturas ramificadas, semelhantes a alambres, chamadas de axônios e dendritos. Quando o axônio de um neurônio toca o dendrito de outro, são formadas pequenas conexões chamadas sinapses. É nas sinapses que o impulso nervoso de uma célula influi no comportamento de outra célula. Se chega um pico a uma sinapse, é possível que também chegue à célula receptora. Algumas sinapses têm o efeito contrário e tornam menos provável que a célula receptora também gere um pico. A força de uma sinapse pode mudar de acordo com o comportamento das duas células. A forma mais simples desta mudança sináptica é o aumento da força de conexão entre dois neurônios quando ambos geram um pico quase ao mesmo tempo. Analisarei mais este processo, chamado de aprendizagem hebbiana, um pouco mais adiante. Além de mudar a força de uma sinapse, há provas que indicam que podem ser formadas sinapses completamente novas entre dois neurônios. Talvez isto aconteça de forma contínua, ainda que as provas científicas sejam polêmicas. Deixando de lado os detalhes sobre como as sinapses mudam suas forças, o que realmente é certo é que a formação e a força das sinapses é o que faz com que as memórias se armazenem.

Ainda que haja muitos tipos de neurônios no neocórtex, uma ampla classe compreende oito de cada dez células. Trata-se dos neurônios piramidais, assim chamados porque seus corpos celulares apresentam uma forma parecida às pirâmides. Salvo a camada superior das seis que formam o córtex, que tem

milhares de axônios, mas muito poucas células, as restantes contêm células piramidais. Cada neurônio piramidal liga-se com muitos outros das suas imediações, e cada um envia um longo axônio lateral a regiões mais distantes do córtex ou a estruturas cerebrais inferiores como o tálamo.

Uma célula piramidal comum possui várias milhares de sinapses. Mais uma vez, acaba sendo muito difícil saber com exatidão quantas devido a sua extrema densidade e reduzido tamanho. O número de sinapses varia de uma célula a outra, de uma camada a outra e de uma região a outra. Se aceitamos a posição conservadora de que a célula piramidal média tem mil sinapses (é provável que o número real se aproxime à cinco ou à dez mil), nosso neocórtex se aproximaria à trinta trilhões de sinapses. É um número astronômico que vai além do nosso entendimento intuitivo. Parece ser suficiente para guardar todas as coisas que você pode aprender em uma vida.

* * *

Segundo rumores, Albert Einstein afirmou certa vez que conceber a teoria da relatividade especial tinha sido algo imediato, quase fácil. Deduziu-a de forma natural em uma única observação: a velocidade da luz é constante para todos os observadores, ainda que estes se movam a velocidades diferentes. Isto vai contra a intuição. É como dizer que a velocidade de uma bola lançada é sempre a mesma independentemente da força com a qual se lance ou da rapidez com que se movam os indivíduos que a lançam e observam. Todos veem a bola se movendo à mesma velocidade em relação a eles em todas as circunstâncias. Não parece que isto possa ser verdade, mas se demonstrou que era assim pela luz; e o inteligente Einstein perguntou-se quais eram as consequências deste estranho fato. Ele pensou metodicamente em todas as repercussões de uma velocidade de luz constante, o que lhe conduziu às predições ainda mais estranhas da relatividade especial, tais como a de que o tempo andava mais devagar quando você avançava mais rápido, e que a energia e a massa eram em essência o mesmo. Os livros sobre a relatividade percorrem esta linha de raciocínio com exemplos quotidianos de trens, balas, lanternas e assim por diante. A teoria não é difícil, mas sem dúvida vai contra a intuição.

Existe uma descoberta análoga na neurociência, um fato sobre o córtex que acaba sendo tão surpreendente que alguns neurocientistas se negam a crer e a maioria restante o passa por alto porque não sabe o que fazer com ele. Mas é um fato de tal importância que se forem exploradas suas consequências cuidadosa e metodicamente desvelará os segredos do que faz o neocórtex e como funciona. Neste caso, a descoberta surpreendente provém da anatomia básica do próprio córtex, mas foi preciso uma mente com uma perspicácia fora do comum para reconhecer isto. Essa pessoa foi Vernon Mountcastle, neurocientista da Universidade John Hopkins de Baltimore. Em 1978, ele publicou um artigo titulado "An Organizing Principle for Cerebral Function", no qual ele assinala que o

neocórtex é notavelmente uniforme quanto à aparência e estrutura. As regiões que se ocupam das entradas auditivas se assemelham às que se ocupam do tato, que se parecem às regiões que controlam os músculos, semelhantes à área de linguagem de Broca, que é parecida a quase todas as regiões restantes do córtex. Mountcastle sugere que já que as ditas regiões parecem semelhantes, talvez realizem a mesma operação básica. Ele propõe que o córtex usa a mesma ferramenta computacional para realizar tudo o que faz.

Os anatomistas da época e das décadas anteriores a Mountcastle reconheciam que o córtex era semelhante em todas as suas partes; isto era algo inegável. Mas em vez de perguntar-se o que isto podia significar, dedicaram seu tempo a buscar diferenças entre uma área e outra. E encontraram-nas. Assumiram que se uma região se emprega para a linguagem e outra para a visão, deveria haver diferenças entre ambas. Se você as busca com o cuidado suficiente, você as encontra. As regiões do córtex variam em espessura, densidade celular, proporção relativa de células e muitos outros aspectos que podem acabar sendo difíceis de descobrir. Uma das regiões mais estudadas, a área visual primária V1, apresenta mais algumas divisões em uma das suas camadas. A situação é análoga ao trabalho dos biólogos em meados da década de 1800. Eles dedicaram seu tempo a descobrir as diferenças mínimas entre as espécies. O sucesso consistia em descobrir que dois ratos que pareciam quase idênticos eram na realidade espécies separadas. Durante muitos anos, Darwin seguiu o mesmo curso, estudando com frequência moluscos. Mas acabou tendo a grande percepção de perguntar-se por que todas essas espécies podiam ser tão parecidas. É sua semelhança que acaba sendo surpreendente e interessante, bem mais do que suas diferenças.

Mountcastle realizou uma observação similar. Em um campo de anatomistas que buscam diferenças mínimas nas regiões corticais, ele mostrou que, apesar das diferenças, o neocórtex é notavelmente uniforme. As mesmas camadas, tipos de células e conexões existem por todas as partes. Todas elas são parecidas com os seis cartões de visita. As diferenças são com frequência tão sutis que nem os anatomistas experts conseguem chegar a um acordo a respeito. Portanto, sustenta Mountcastle, todas as regiões do córtex executam as mesmas operações. O que faz com que a área da visão seja visual e a área do movimento seja motora é o modo como as diversas regiões do córtex estão ligadas entre si e com outras partes do sistema nervoso central.

De fato, Mountcastle sustenta que a razão de uma região do córtex parecer ligeiramente diferente de outra são as suas conexões, e não a sua função básica. Ele conclui que existe uma função comum, um algoritmo comum que todas as regiões corticais o executam. A visão não é diferente da audição, que não é diferente de uma saída motora. Ele permite que nossos genes especificam como serão ligadas as regiões, o que é muito peculiar da função e da espécie, mas o tecido cortical em si faz o mesmo em todos os lugares.

Pensemos sobre isso um momento. Para mim, a visão, a audição e o tato parecem muito diferentes. Possuem qualidades essencialmente diferentes. A visão envolve cor, textura, contorno, profundidade e forma. A audição tem tom, ritmo e timbre. Parecem muito diferentes. Como podem ser o mesmo? Mountcastle afirma que não são, mas o modo como o córtex processa os sinais procedentes do ouvido é o mesmo que ele emprega para processar os sinais dos olhos. Ele prossegue dizendo que o controle motor funciona também segundo o mesmo princípio.

Em sua maior parte, os cientistas e engenheiros têm ignorado a proposta de Mountcastle, ou têm preferido passá-la por alto. Quando tentam entender a visão ou fabricar um computador capaz de “ver”, eles criam vocabulário e técnicas específicos para a visão. Falam de bordas, texturas e representações tridimensionais. Se eles querem compreender a linguagem falada, constroem algoritmos baseados em regras gramaticais, sintaxe e semântica. Mas se Mountcastle estiver certo, as ditas propostas não se ajustam ao modo como o cérebro resolve estes problemas, portanto, é provável que fracassem. Se Mountcastle estiver com a razão, o algoritmo do córtex deve ser expressado com independência de qualquer função ou sentido particular. O cérebro emprega o mesmo processo tanto para ver como para ouvir. O córtex faz algo universal que pode ser aplicado a qualquer tipo de sistema sensorial ou motor.

Quando li pela primeira vez o artigo de Mountcastle, quase caí da cadeira. Ali estava a pedra Roseta da neurociência, um único artigo e uma única ideia que unia todas as faculdades distintas e maravilhosas da mente humana. Ele as unia sob um único algoritmo. Com um único passo ele deixava exposta a falácia de todas as tentativas anteriores de entender e engenhar o comportamento humano com suas capacidades diversas. Espero que você seja capaz de apreciar a elegância radical e maravilhosa da proposta de Mountcastle. As melhores ideias da ciência sempre são simples, elegantes e inesperadas, e esta é uma das melhores. Na minha opinião, foi, é, e provavelmente continuará sendo a descoberta mais importante da neurociência. No entanto, por incrível que pareça, a maioria dos cientistas e engenheiros se negam a crer nela, preferem a passar por alto ou não a conhecem.

* * *

Parte desta negligência é devida à escassez de ferramentas para estudar como flui a informação dentro das seis camadas do córtex. As ferramentas com que contamos operam a um nível grosseiro e em geral seu objetivo consiste em determinar onde —ao invés de quando e como— surgem as diversas faculdades no córtex. Por exemplo, boa parte da neurociência que aparece na imprensa popular dos nossos dias favorece de forma implícita a ideia de que o cérebro é uma reunião de módulos de alta especialização. As técnicas de imagem funcional como os escâneres funcionais MRI e PET se centram quase com exclusividade nos mapas cerebrais e nas regiões funcionais que já mencionei. Nestes experimentos, um sujeito voluntário fica deitado com a cabeça dentro do escâner e executa um tipo

de tarefa mental ou motora. Poderia estar jogando um videogame, gerando conjugações verbais, lendo orações, olhando rostos, descrevendo fotos, imaginando algo, memorizando listas, tomando decisões financeiras e assim por diante. O escâner detecta quais regiões do cérebro estão mais ativas do que o habitual durante estas tarefas e desenha manchas coloridas sobre uma imagem do cérebro do sujeito para indicá-las. Ao que parece, estas regiões são cruciais para a tarefa. Realizaram-se milhares de experimentos de imagens funcionais, e continuarão sendo realizados mais milhares. Com todos eles estamos construindo pouco a pouco um quadro indicador de onde acontecem certas funções no cérebro adulto normal. É fácil afirmar: "esta é a área de reconhecimento de rostos, esta é a área da matemática, esta é a área da música", e assim sucessivamente. Como não sabemos como executa o cérebro as ditas tarefas, acaba sendo natural assumir que o faz de modos diferentes.

Mas será que é isso mesmo? Um conjunto crescente e fascinante de provas apoia a proposta de Mountcastle. Alguns dos melhores exemplos demonstram a extrema flexibilidade e plasticidade do neocórtex. Todo cérebro humano, se for nutrido adequadamente e se for colocado no ambiente preciso, poderá aprender qualquer uma das milhares de línguas faladas. Esse mesmo cérebro também será capaz de aprender a linguagem dos sinais, a linguagem escrita, a linguagem musical, a linguagem matemática, as linguagens computacionais e a linguagem corporal. Poderá aprender a viver nos gelados climas do Norte ou em um deserto abrasador. Poderá chegar a ser um expert em xadrez, pesca, agricultura ou física teórica. Consideremos o fato de que temos uma pequena área visual especial que parece ser dedicada a representar letras e dígitos escritos. Significa isto que nascemos com uma área de linguagem pronta para processar letras e dígitos? Não é muito provável. A linguagem escrita é um invento muito recente para que nossos genes possam ter evoluído um mecanismo específico a respeito. Portanto, o córtex ainda continua se dividindo em áreas funcionais com tarefas específicas até avançar a infância, baseadas puramente na experiência. O cérebro humano possui uma capacidade incrível de aprender e adaptar-se a milhares de ambientes que não existiam até data muito recente. Isto argumenta a favor de um sistema com uma flexibilidade incrível, não de um com milhares de soluções para milhares de problemas.

Os neurocientistas também descobriram que o sistema de conexões do neocórtex é surpreendentemente "plástico", o que significa que ele pode mudar e reconectar-se segundo o tipo de entradas que receber. Por exemplo, as conexões nos furões recém nascidos podem ser mudadas cirurgicamente para que os olhos do animal enviem seus sinais às áreas do córtex onde normalmente se desenvolve a audição. O resultado surpreendente é que são desenvolvidos caminhos visuais funcionais nas porções auditivas dos seus cérebros. Em outras palavras, veem com o tecido cerebral que normalmente escuta sons. Efetuaram-se experimentos similares com outros sentidos e regiões cerebrais. Por exemplo, logo ao nascer, pode-se

transplantar pedaços de córtex visual de um rato para regiões nas quais se costuma representar o sentido do tato. Quando o rato cresce, o tecido transplantado processa o tato em vez da visão. As células não nascem para se especializar em visão, tato ou audição.

O neocórtex humano é muito plástico. Os adultos que nascem surdos processam a informação visual em áreas que normalmente se tornam regiões auditivas. E os adultos cegos de nascimento usam a parte mais atrás do seu córtex, que em geral se dedica à visão, para ler Braille. Como o Braille envolve tato, caberia pensar que a princípio se ativaria as regiões do tato, mas ao que parece nenhuma área do córtex se contenta em não representar nada. O córtex visual, ao não receber informação dos olhos como se “supõe”, tenta encontrar ao redor outros padrões de entrada para passar por eles; neste caso, de outras regiões corticais.

Tudo isto pretende mostrar que as regiões cerebrais desenvolvem funções especializadas baseadas em boa medida no tipo de informação que flui a elas durante o desenvolvimento. O córtex não está projetado de forma rígida para realizar diferentes funções utilizando diferentes algoritmos, do mesmo modo que a superfície da Terra não estava predestinada a acabar no seu moderno ordenamento de nações. A organização do nosso córtex, assim como a geografia política do globo, poderia ter acabado sendo diferente se tivesse sido dado um conjunto de circunstâncias diferentes.

Os genes ditam a arquitetura geral do córtex, incluindo as especificações de quais regiões serão ligadas entre si, mas dentro dessa estrutura o sistema é muito flexível.

Mountcastle estava certo. Não há mais do que um único algoritmo posto em prática por cada uma das regiões do cérebro. Se as regiões do córtex forem ligadas em uma hierarquia apropriada e lhes forem proporcionado um fluxo de entrada, elas aprenderão do seu ambiente. Portanto, não há razão para que as máquinas inteligentes do futuro tenham os mesmos sentidos ou faculdades que nós humanos. O algoritmo cortical pode ser utilizado de maneiras inovadoras, com novos sentidos, em uma lâmina cortical fabricada de maneira que surja uma inteligência real e flexível fora dos cérebros biológicos.

* * *

Passemos agora a um tema que está relacionado com a proposta de Mountcastle e que é muito surpreendente. As entradas que nosso córtex recebe são todas basicamente iguais. Mais uma vez, é provável que você pense que seus sentidos são entidades bem separadas. Além disso, o som se transporta como ondas de compressão pelo ar; a visão, como luz, e o tato, como uma pressão sobre a pele. O som parece temporal; a visão, sobretudo pictorial; e o tato, espacial. O que

poderia ser mais diferente do que o som de uma cabra berrando versus a visão de uma maçã versus o tato de uma bola de beisebol?

Mas observemos com maior detalhamento. A informação visual procedente do mundo exterior é enviada ao nosso cérebro através de um milhão de fibras do nervo óptico. Após um breve trânsito pelo tálamo, ela chega ao córtex visual primário. Os sons são enviados através de trinta mil fibras do nervo auditivo. Passam por algumas partes mais antigas do cérebro e depois chegam ao córtex auditivo primário. A medula espinhal transfere informação do tato e das sensações internas para o cérebro através de outro milhão de fibras, que são recebidas pelo córtex somatosensorial primário. Estas são as principais entradas do nosso cérebro. São como sentimos o mundo.

Cabe visualizar estas entradas como um pacote de cabos elétricos ou um punhado de fibras ópticas. Talvez você já tenha visto lustres feitos com fibras ópticas onde aparecem pontos de luz colorida no final de cada uma. As entradas ao cérebro são semelhantes, mas as fibras são chamadas de axônios e transportam sinais neurais chamados de “potenciais de ação” ou “picos”, que são em parte químicos e em parte elétricos. Os órgãos sensoriais que fornecem os ditos sinais são diferentes, mas uma vez que se convertem em potenciais de ação dirigidos ao cérebro se tornam tudo a mesma coisa: nada mais do que padrões.

Se olharmos um cão, por exemplo, um conjunto de padrões fluirá pelas fibras do nosso nervo óptico até a parte visual do córtex. Se escutarmos o cão latir, fluirá um conjunto diferente de padrões por nosso nervo auditivo até as partes auditivas do cérebro. Se acariciarmos o cão, um conjunto de padrões de tato-sensação fluirá da nossa mão através das fibras da medula espinhal até as partes do cérebro que se ocupam do tato. Cada padrão —ver o cão, escutar o cão, sentir o cão— é experimentado de forma diferente porque cada um se canaliza por um caminho diferente na hierarquia cortical. É importante por onde entram no cérebro os cabos. Mas no nível abstrato das entradas sensoriais são todos em essência o mesmo, e todos se manejam de forma similar pelas seis camadas do córtex. Escutamos o som, vemos a luz e sentimos a pressão, mas dentro do nosso cérebro não existe nenhuma diferença fundamental entre esses tipos de informação. Uma ação potencial é uma ação potencial. Estes picos momentâneos são idênticos, independentemente do que originariamente os causou. Todo nosso cérebro conhece estes padrões.

Nossas percepções e conhecimento sobre o mundo são construídos com estes padrões. Não há luz dentro das nossas cabeças; há escuridão. Também não entra som no cérebro; dentro há silêncio. De fato, o cérebro é a única parte do nosso corpo que não tem sentidos. Um cirurgião poderia cravar um dedo dentro do cérebro e não o sentiríamos. Toda a informação que entra em nossa mente chega como padrões espaciais e temporais nos axônios.

O que entendo exatamente por padrões espaciais e temporais? Observemos cada um dos nossos principais sentidos. A visão transporta informação espacial e temporal. Os *padrões espaciais* são padrões coincidentes no tempo; criam-se quando múltiplos receptores do mesmo órgão sensorial se estimulam de forma simultânea. Na visão, o órgão sensorial é a retina. Quando entra uma imagem na pupila, a mesma é invertida pelas lentes, atinge a retina e é criado um padrão espacial. Este padrão é enviado ao cérebro. A gente tende a pensar que é uma pequena foto invertida do mundo que vai às áreas visuais, mas não é bem assim o que acontece. Não há nenhuma foto. Ela deixou de ser uma imagem. Em essência, não é mais do que atividade elétrica emitindo padrões. Suas qualidades de imagem são perdidas rapidamente quando o córtex maneja a informação, passando os componentes do padrão acima e abaixo entre as diferentes áreas, os mudando e os filtrando.

A visão também se baseia em padrões temporais, o que significa que os padrões que entram nos olhos mudam constantemente ao longo do tempo. Mas enquanto o aspecto espacial da visão pareça óbvio por intuição, seu aspecto temporal é menos evidente. Uma três vezes por segundo os olhos fazem um movimento repentino chamado de sacada. Eles fixam-se em um ponto e depois de improviso saltam a outro. Cada vez que os olhos se movem, a imagem da retina muda, o que significa que os padrões transportados ao cérebro também mudam por completo com cada sacada ocular. E assim ocorre no caso mais simples possível, mesmo quando olhamos sentados uma cena que não muda. Na vida real, movemos constantemente a cabeça e o corpo, e caminhamos por ambientes que variam de forma contínua. Nossa impressão consciente é de que há um mundo estável cheio de objetos e pessoas que parece ser fácil seguir. Mas esta impressão só é possível devido a nossa capacidade cerebral de manejar uma torrente de imagens retinianas que nunca repetem um padrão exato. A visão natural, experimentada como padrões que entram no cérebro, flui como um rio. A visão parece-se mais com uma canção do que com uma pintura.

Muitos pesquisadores da visão ignoram as sacadas oculares e os padrões em mudança constante da visão. Trabalhando com animais anestesiados, eles estudam como atua a visão quando um animal inconsciente a fixa em um ponto. Ao fazer isto, estão desprezando a dimensão temporal. Em princípio não há nada de mau; eliminar variáveis é um elemento central do método científico. Mas estão eliminando um componente crucial da visão, o que realmente a constitui. O tempo deve ocupar um lugar central em uma explicação neurocientífica da visão.

Quanto à audição, estamos acostumados a pensar no aspecto temporal do som. É evidente por intuição que os sons, a linguagem falada e a música mudam com o tempo. Não se pode escutar uma canção completa ao mesmo tempo, do mesmo modo que não é possível ouvir uma oração falada em apenas um instante. Uma canção só existe ao longo do tempo. Mas nem sempre pensamos nos sons como

um modelo espacial. De certa forma, é o contrário do caso da visão: o aspecto temporal parece evidente de imediato, mas seu aspecto espacial é menos óbvio.

A audição também tem um componente espacial. Convertemos os sons em potenciais de ações mediante um órgão enrolado em cada orelha chamado de cóclea. Diminuto, opaco, com forma de espiral e inserido no osso mais duro do corpo, o osso temporal, a cóclea, foi decifrada há mais de meio século por um físico húngaro, Georg von Beksey. Construindo modelos do ouvido interno, Von Beksey descobriu que cada componente do som que ouvimos faz com que vibre uma parte diferente da cóclea. Os tons de frequência alta provocam vibrações na base rígida da cóclea. Os tons de frequência baixa causam vibrações em sua parte mais flexível e larga. Os tons de frequência média fazem vibrar os segmentos intermediários. Cada lugar da cóclea está salpicado de neurônios que se estimulam quando são agitados. Na vida cotidiana, a nossa cóclea está sendo vibrada por grandes quantidades de frequências simultâneas o tempo todo. Portanto, a cada momento há um novo padrão espacial de estimulação por toda a extensão da cóclea; a cada momento um novo padrão espacial flui até o nervo auditivo. Novamente, vemos que esta informação sensorial se converte em padrões espaço-temporais.

A gente não costuma pensar que o tato é um fenômeno temporal, mas está tão baseado no tempo como no espaço. Você pode efetuar um experimento para comprovar isto. Peça a um amigo para que ele levante a mão com a palma para cima e feche os olhos. Coloque um pequeno objeto comum na palma dele —um anel, uma borracha, qualquer coisa servirá— e peça para que ele o identifique sem mover nenhuma parte da mão. Ele não terá mais pista do que o peso e talvez o tamanho bruto. Depois peça para que ele mantenha os olhos fechados e mova os dedos sobre o objeto. É muito provável que ele o identifique de imediato. Ao permitir que os dedos se movam, você acrescentou tempo à percepção sensorial do tato. Existe uma analogia direta entre a fóvea do centro da retina e as pontas dos dedos, pois ambas possuem uma grande precisão. Portanto, o tato também é como uma canção. Nossa capacidade para fazer um uso complexo do tato, como abotoarmos a camisa ou abrir o trinco da porta da frente na escuridão, depende de padrões do sentido do tato que variam constantemente com o tempo.

Ensinamos a nossos filhos que os humanos gozam de cinco sentidos: visão, audição, tato, olfato e paladar, mas na realidade temos mais. A visão é mais três sentidos: movimento, cor e luminância (contraste de preto-e-branco). O tato tem pressão, temperatura, dor e vibração. Também contamos com um sistema completo de sensores que nos informam sobre nossos ângulos de união e posição corporal. Chama-se sistema proprioceptivo (*próprio* tem a mesma raiz latina que *proprietário* e *propriedade*). Não poderíamos nos mover sem ele. Também dispomos do sistema vestibular do ouvido interno, que nos proporciona o sentido do equilíbrio. Alguns destes sentidos são mais ricos e evidentes que outros, mas

todos entram em nosso cérebro como uma corrente de padrões espaciais que fluem através do tempo nos axônios.

Na realidade, nosso córtex não conhece nem sente o mundo de forma direta. A única coisa que conhece é o padrão que chega nos axônios de entrada. Nossa visão percebida do mundo é criada com estes padrões, incluindo nosso sentimento de nós mesmos. De fato, nosso cérebro não pode conhecer de forma direta onde termina nosso corpo e começa o mundo. Os neurocientistas que estudam a imagem corporal têm descoberto que nosso sentido do *eu* é bem mais flexível do que parece. Por exemplo, se lhe dou um rastelo pequeno e lhe digo que o use para atingir e apanhar coisas em vez de empregar a mão, em breve você sentirá que ele se converteu em parte do seu corpo. Seu cérebro mudará suas expectativas para acomodar-se aos novos padrões de entrada tátil. O rastelo é incorporado literalmente ao seu mapa corporal.

* * *

A ideia de que os padrões de diferentes sentidos são equivalentes dentro do nosso cérebro é bastante surpreendente e, ainda que se entenda bem, continua sem se apreciar por completo. Temos mais exemplos. O primeiro pode ser reproduzido em casa. Tudo o que se precisa é de um amigo, uma tela de cartolina e uma mão de mentira. Para realizar pela primeira vez este experimento seria ideal contar com uma mão de borracha das que se podem comprar em uma loja de brinquedos, mas também serviria uma mão pintada em uma folha branca de papel. Coloque sua mão real sobre uma mesa a alguns centímetros da falsa e situe-as do mesmo modo (com as pontas dos dedos apontando na mesma direção e ambas as palmas para cima ou para baixo). Depois ponha a tela entre as duas mãos de maneira que você só consiga ver a mão falsa. Enquanto você olha fixamente a mão falsa, a tarefa do seu amigo consistirá em golpear ambas as mãos em pontos correspondentes. Por exemplo, seu amigo pode golpear ambos midinhos desde o dedo até a unha na mesma velocidade, depois dar três rápidos toques na segunda articulação de ambos os dedos indicadores com o mesmo ritmo, a seguir desenhar alguns círculos ligeiros no dorso de cada mão, e assim sucessivamente. Decorrido algum tempo, as áreas do seu cérebro em que chegam juntos os padrões visuais e somatosensoriais —uma das áreas de associação que já mencionei neste capítulo— acabam sendo confundidas. Você sentirá as sensações correspondentes à mão de borracha como se fossem suas.

Outro exemplo fascinante desta “equivalência de padrões” chama-se substituição sensorial. Ela pode revolucionar a vida de gente que perdeu a visão na infância, e talvez em algum dia possa ser de grande ajuda para as pessoas que nasceram cegas. Ela também poderia produzir novas tecnologias de interface para fabricar máquinas úteis para o resto de nós.

Dando-se conta de que no cérebro tudo consiste em padrões, Paul Bach e Rita, professor de engenharia biomédica na Universidade de Wisconsin, desenvolveram um método para se detectar padrões visuais na língua humana. Usando um aparelho de visualização, as pessoas cegas estão aprendendo a “ver” através das sensações da língua.

Assim é como funciona. O sujeito leva uma pequena câmera na frente. As imagens visuais transferem-se pixel a pixel à pontos de pressão na língua. Uma cena visual que se pode traduzir como centenas de pixels sobre uma tela de televisão comum se converte em um padrão de diminutos pontos de pressão sobre a língua. O cérebro aprende em seguida a interpretar bem os padrões.

Uma das primeiras pessoas a levar o aparelho montado na língua é Erik Weihenmayer, atleta de categoria mundial que ficou cego aos treze anos e que dá muitas conferências sustentando que não vai permitir que a cegueira detenha suas ambições. Em 2002, Weihenmayer escalou o Everest e converteu-se na primeira pessoa cega que tinha tentado, e muito menos atingido, essa meta.

Em 2003, ele provou a unidade colocada na língua e viu imagens pela primeira vez desde sua infância. Ele foi capaz de distinguir uma bola que rolava no chão para ele, apanhar um refresco da mesa e jogar “pedra, papel, tesoura”. Depois ele caminhou por um corredor, viu a abertura das portas, examinou uma porta e seu modelo, e notou que existia uma indicação nela. As imagens experimentadas no início como sensações na língua rapidamente passaram a ser percebidas como imagens no espaço.

Estes exemplos mostram mais uma vez que o córtex é extremamente flexível e que as entradas que chegam ao cérebro são simplesmente padrões. Não importa de onde chegam estes; sempre que tiverem uma correlação temporal coerente, o cérebro poderá achar sentido nos mesmos.

* * *

Tudo isto não deve parecer muito surpreendente se adotamos a proposta de que o cérebro só conhece padrões. Os cérebros são máquinas de padrões. Não é errôneo expressar as funções cerebrais em termos de audição ou visão, mas no nível mais fundamental os padrões são a essência do jogo. Por mais diferentes que possam parecer as atividades de várias áreas corticais, nelas funciona o mesmo algoritmo cortical básico. O córtex não se importa se os padrões foram originados na visão, na audição ou em outro sentido. Não se importa se suas entradas chegam de um único órgão sensorial ou de quatro. Nem se importaria também se acontecesse a casualidade de percebermos o mundo com sonar, radar ou campos magnéticos, ou se tivéssemos tentáculos em vez de mãos, ou inclusive se vivêssemos em um mundo de quatro dimensões em vez de três.

Isso significa que não precisamos de nenhum dos sentidos ou nenhuma combinação particular de sentidos para sermos inteligentes. Helen Keller não tinha nem visão nem audição, mas ela aprendeu a linguagem e se converteu em uma escritora mais capacitada do que a maioria das pessoas que veem e ouvem. Ela era uma pessoa muito inteligente sem os dois dos nossos principais sentidos, mas a incrível flexibilidade do cérebro permitiu-lhe perceber e compreender o mundo assim como fazem as pessoas com os cinco sentidos.

Este tipo de flexibilidade notável na mente humana proporciona-me grandes esperanças a respeito da tecnologia baseada no cérebro que criaremos. Quando penso em construir máquinas inteligentes, me pergunto por que deveríamos nos limitar a nossos sentidos conhecidos. Assim que pudermos decifrar o algoritmo neocortical e elaborar uma ciência sobre os padrões, poderemos aplicar isto a qualquer sistema que queiramos fazer inteligente. E uma das grandes características do sistema de circuitos inspirado no córtex é que não precisaremos ser especialmente inteligentes para o programar. Do mesmo modo que o córtex auditivo pode se converter em "visual" em um furão reconectado, e do mesmo modo que o córtex visual encontra um uso alternativo nas pessoas cegas, um sistema que leve o algoritmo neocortical será inteligente baseado em qualquer tipo de padrões que decidamos lhe dar. No entanto, ainda sim é preciso que sejamos inteligentes para organizar os amplos parâmetros do sistema, pois será necessário o treinar e o educar. Mas os bilhões de detalhes neurais que tomam parte na capacidade do cérebro de ter pensamentos complexos e criativos se ocuparão de si mesmos de forma tão natural assim como fazem em nossas crianças.

Por último, a ideia de que os padrões são a moeda fundamental da inteligência conduz a algumas questões filosóficas interessantes. Quando me sento em uma sala com meus amigos, como sei que eles estão lá, ou inclusive se eles são reais? Meu cérebro percebe um conjunto de padrões que são consistentes com outros que experimentei no passado. Estes padrões correspondem às pessoas que conheço, seus rostos, suas vozes, seu comportamento habitual, e todo tipo de dados sobre elas. Aprendi a esperar que estes padrões ocorram juntos de formas previsíveis. Mas quando você chega a isso, não se trata mais do que um modelo. Todo nosso conhecimento do mundo é um modelo baseado em padrões. Estamos seguros de que o mundo é real? Parece divertido e estranho pensar nisso. Vários livros e filmes de ficção científica exploram este tema. Não se trata de afirmar que as pessoas ou os objetos não estejam lá. Estão realmente lá. Mas nossa certeza da existência do mundo baseia-se na coerência dos padrões e em como os interpretamos. A percepção direta não existe. Não temos um sensor de "pessoas". Recordemos que o cérebro está em uma caixa escura e silenciosa, sem nenhum conhecimento além dos padrões que fluem ao longo do tempo em suas fibras de entrada. Nossa percepção do mundo é criada a partir desses padrões, nada mais. A existência pode ser objetiva, mas os padrões espaço-temporais que fluem nos feixes de axônios do nosso cérebro é tudo o que temos para seguir adiante.

Esta discussão ressalta a relação às vezes questionada entre alucinação e realidade. Se pode-se alucinar sensações provenientes de uma mão de borracha e pode-se “ver” através da estimulação do tato da língua, estamos sendo igualmente “enganados” quando sentimos o tato em nossa própria mão ou vemos com nossos olhos? Podemos confiar em que o mundo é como parece? Sim. O mundo existe de uma forma absoluta muito próxima a como o percebemos. No entanto, nossos cérebros não podem conhecer o mundo absoluto de modo direto.

O cérebro sabe do mundo através de um conjunto de sentidos que só podem detectar partes do mundo absoluto. Os sentidos criam padrões que são enviados ao córtex e processados pelo mesmo algoritmo cortical para criar um modelo do mundo. Deste modo, a linguagem falada e a linguagem escrita percebem-se de forma muito similar, apesar de serem completamente diferentes a nível sensorial. Mesmo assim, o modelo de Helen Keller do mundo estava muito próximo do seu e do meu, apesar do fato de que ela possuía um conjunto de sentidos muito reduzido. Mediante estes padrões o córtex constrói um modelo do mundo que se aproxima à coisa real, e depois, surpreendentemente, o memoriza. A memória, o que acontece a estes padrões uma vez que entram no córtex, será o tema do próximo capítulo.

A Memória

Quando você lê este livro, caminha por uma rua cheia de gente, escuta uma sinfonia ou consola uma criança que chora, seu cérebro é inundado com os padrões espaciais e temporais provenientes de todos seus sentidos. O mundo é um oceano de padrões em mudança constante que se agita e se choca contra os nossos cérebros. Como conseguimos achar sentido nessa avalanche? Os padrões chegam, passam por várias partes do cérebro antigo e acabam no neocórtex. Mas o que lhes acontece quando entram nele?

Desde o surgimento da Revolução Industrial, pessoas têm visto o cérebro como a uma espécie de máquina. Sabiam que ele não tinha engrenagens nem dentes, mas era a melhor metáfora de que dispunham. A informação entrava no cérebro de algum modo e a máquina-cérebro determinava como devia reagir o corpo. Durante a era da informática, o cérebro foi considerado um tipo de máquina particular, o computador programável. E, como temos visto no primeiro capítulo, os pesquisadores da inteligência artificial têm ficado presos a esta postura, sustentando que a sua falta de avanço só se deve ao fato dos computadores continuarem sendo pequenos e lentos em comparação com o cérebro humano. Os computadores atuais só equivalem ao cérebro de um crocodilo, afirmam, mas quando os fabricarmos maiores e mais rápidos serão inteligentes como os humanos.

Nesta analogia do cérebro como um computador há um problema ignorado em boa medida. Os neurônios são bastante lentos comparados com os transistores de um computador. Um neurônio reúne entradas das suas sinapses e combina-as para decidir quando enviar um impulso a outro neurônio. Um neurônio normal pode fazer isto e se resetar em uns cinco milésimos de segundo, ou cerca de duzentas vezes por segundo. Talvez pareça rápido, mas um computador moderno de silício pode realizar bilhões de operações em um segundo, o que significa que uma operação computacional básica é cinco milhões de vezes mais rápida que a operação mais elementar do nosso cérebro. Trata-se de uma diferença gigantesca. De modo que, como é possível que um cérebro possa ser mais rápido e mais potente do que nossos computadores digitais mais velozes? “Não há problema” — dizem as pessoas que opinam que o cérebro é um computador—. “O cérebro é um computador paralelo. Possui bilhões de células computando todas ao mesmo tempo. Este paralelismo multiplica significativamente o poder de processamento do cérebro humano.”

Sempre me pareceu que este argumento é uma falácia, e um simples experimento mental mostra por que. Chama-se a “regra dos cem passos”. Um humano pode realizar tarefas consideráveis em muito menos tempo que um segundo. Por exemplo, eu poderia te mostrar uma fotografia e pedir que indicasse se há um

gato na imagem. Sua tarefa seria pulsar um botão se visse um gato, mas não o fazer se visse um urso, um javali ou um nabo. Esta tarefa é difícil ou impossível de realizar para um computador atual, mas um humano pode fazer isto de forma confiável em meio segundo ou menos. Mas os neurônios são lentos, de modo que nesse meio segundo a informação que entra em nosso cérebro só é capaz de atravessar uma corrente de cem neurônios. Isto é, o cérebro “computa” soluções a problemas em cem passos ou menos, independentemente de quantos neurônios possam participar ao todo. Desde o instante em que a luz entra em nosso olho até o momento em que pulsamos o botão, poderia participar uma corrente não mais longa do que cem neurônios. Um computador digital que tentasse resolver o mesmo problema precisaria de bilhões de passos. Cem instruções computacionais mal bastam para mover um único caráter na tela do computador, que diria para fazer algo interessante.

Mas se tenho muitos milhares de neurônios trabalhando juntos, não se pareceria a um computador paralelo? Não. Os cérebros operam em paralelo e os computadores paralelos operam em paralelo, mas é a única coisa que eles têm em comum. Computadores paralelos são combinados com muitos outros computadores rápidos para trabalhar em grandes problemas, como calcular a previsão do tempo do dia seguinte. Para predizer o tempo têm-se que computar as condições físicas em muitos pontos do planeta. Cada computador pode trabalhar em uma localização diferente ao mesmo tempo. Mas ainda que seja possível ter centenas ou inclusive milhares de computadores trabalhando em paralelo, cada um deles em particular continua precisando realizar bilhões ou trilhões de passos para executar suas tarefas. O computador paralelo concebível não é capaz de fazer nada útil em cem passos por mais grande ou rápido que seja.

Vejamos uma analogia. Suponhamos que eu te peça para transportar cem blocos de pedra para o outro lado de um deserto. Você poderá levar as pedras uma à uma, o que precisará de um milhão de passos para cruzar o deserto. Você se dá conta de que demorará muito em conseguir sozinho, de modo que contrata cem trabalhadores para que façam isto em paralelo. A tarefa agora avança cem vezes mais rápido, mas continua requerendo no mínimo de um milhão de passos para cruzar o deserto. Contratar mais trabalhadores—inclusive mil— não proporcionaria uma vantagem adicional. Por mais trabalhadores que você disponha, o problema não pode se resolver em menos tempo do que se demora em andar um milhão de passos. O mesmo ocorre no caso dos computadores paralelos. Depois de certo ponto, acrescentar mais processadores não trará nenhuma diferença. Um computador, por mais processadores que tenha e por mais rápido que seja, não pode “computar” a resposta a problemas difíceis em cem passos.

Portanto, como consegue o cérebro realizar tarefas difíceis em cem passos que o maior computador paralelo imaginável não é capaz de realizar em um milhão ou bilhões de passos? A resposta é que o cérebro não “computa” as respostas aos

problemas, mas as recupera da memória. Em essência, as respostas estão armazenadas na memória há um longo tempo. Não é necessário mais do que alguns passos para recuperar algo da memória. Os lentos neurônios não só possuem a rapidez necessária para o fazer, mas eles mesmos constituem a memória. O córtex inteiro é um sistema de memória. Não é de forma alguma um computador.

* * *

Permitam-me mostrar mediante um exemplo a diferença entre *computar* a solução a um problema e *usar a memória* para resolvê-lo. Consideremos a tarefa de apanhar uma bola. Alguém lhe atira uma bola, você a vê se deslocando em sua direção, e em menos de um segundo a pega no ar. Não parece algo muito difícil, até o momento de tentar programar o braço de um robô para que faça isto. Como muitos estudantes graduados têm descoberto, acaba sendo quase impossível. Quando os engenheiros ou cientistas da computação abordam este problema, primeiro eles tentam determinar onde estará a bola quando chegar ao braço. Este cálculo requer resolver um conjunto de equações do tipo das que se aprendem em física no colégio. A seguir eles têm que harmonizar todas as uniões do braço robótico para que movam a mão na posição adequada. Isso envolve resolver outro conjunto de equações matemáticas mais difíceis que as primeiras. Por último, deve se repetir esta operação inteira múltiplas vezes, pois à medida que a bola se aproxima, o robô obtém melhor informação sobre sua localização e trajetória. Se o robô esperar para começar a se mover até conhecer com exatidão onde chegará a bola, será muito tarde para a apanhar. Ele deve começar a avançar para apanhá-la quando mal tem noção da sua localização, e ir se ajustando uma e outra vez enquanto esta se aproxima. Um computador requer milhões de passos para resolver as numerosas equações matemáticas que requer apanhar a bola. E ainda que possa ser programado para solucionar o dito problema, a regra dos cem passos nos indica que um cérebro o resolve de um modo diferente. Ele emprega a memória.

Como se apanha a bola empregando a memória? Nosso cérebro possui uma memória armazenada das ordens musculares requeridas para conseguir isso (junto com muitos outros comportamentos aprendidos). Quando se lança uma bola, ocorrem três coisas. Primeiro, é recuperada de forma automática a memória apropriada diante da visão da bola. Segundo, a memória recorda uma sequência temporal de ordens musculares. E terceiro, a memória recuperada ajusta-se às particularidades do momento, tais como a trajetória presente da bola e a posição do nosso corpo. A memória de como apanhar uma bola não estava programada em nosso cérebro; a aprendemos ao longo de anos de prática repetitiva, e nossos neurônios a guardam, não a calculam.

Talvez pense: "Espere um pouco. Cada ação de apanhar uma bola é ligeiramente diferente. Você acaba de dizer que a memória recuperada tem que se ajustar de

forma contínua para se adaptar às variações da localização em cada lançamento particular... Não requer isso resolver as mesmas equações que estamos tratando de evitar?”. Pode ser que assim pareça, mas a Natureza solucionou o problema da variação de um modo diferente e muito inteligente. Como veremos mais adiante neste mesmo capítulo, o córtex cria o que se chama de *representações invariáveis*, que se ocupam das variações do mundo de forma automática. Uma analogia útil pode ser imaginar no que acontece quando você se senta em uma cama d’água: as almofadas e as outras pessoas da cama são todas empurradas que espontaneamente estabelecem uma nova configuração. A cama não calcula que altura deve se elevar cada objeto; as propriedades físicas da água e o forro plástico do colchão encarregam-se do ajuste de forma automática. Como veremos no próximo capítulo, livremente falando, o projeto do córtex de seis camadas faz algo similar com a informação que flui por ele.

* * *

Assim pois, o neocórtex não se assemelha a um computador, seja paralelo ou de qualquer outro tipo. Em vez de calcular respostas aos problemas, ele utiliza memórias armazenadas para resolver problemas e gerar comportamentos. Os computadores também têm memória na forma de unidades de disco rígido e chips de memória; no entanto, há quatro atributos da memória neocortical que são fundamentalmente diferentes da memória computacional:

- **O neocórtex armazena sequências de padrões.**
- **O neocórtex recorda os padrões por auto-associação.**
- **O neocórtex armazena os padrões em uma forma invariável.**
- **O neocórtex armazena os padrões em uma hierarquia.**

Analisaremos as três primeiras diferenças neste capítulo; no capítulo 3 já apresentei o conceito de hierarquia no neocórtex, e no capítulo 6 descreverei seu significado e funcionamento.

Na próxima vez que você contar uma história, se detenha a pensar sobre o fato de que você só é capaz de narrar um aspecto de vez. Você não pode me contar tudo o que aconteceu ao mesmo tempo, por mais rápido que fale ou eu escute. Você precisa terminar uma parte para passar à próxima. Isso não se deve só porque a linguagem é consecutiva; a narração escrita, oral e visual transmite uma história de forma consecutiva. É assim porque a história se armazena em sua cabeça de forma sequencial e só pode ser recordada na mesma sequência. Não se pode recordar a história inteira de uma só vez. De fato, é quase impossível pensar em algo complexo que não seja uma série de acontecimentos ou pensamentos.

Talvez também se tenha dado conta de que ao contar uma história algumas pessoas não são capazes de chegar ao x da questão. Parecem enfeitar com detalhes mínimos e secundários, o que acaba sendo irritante. Você quer gritar: "Vá direto ao ponto!". Mas elas estão contando a história assim como lhes aconteceu na época e não sabem contar de nenhum outro modo.

Outro exemplo: gostaria que agora você se imaginasse em sua casa. Feche os olhos e visualize-a. Em sua imaginação, vá à porta principal. Imagine seu aspecto. Abra-a. Entre adentro. Agora olhe a sua esquerda. O que vê? Olhe à direita. O que há? Vá ao banheiro. O que há à direita? O que há à esquerda? O que há na gaveta superior direita? Que artigos você guarda na prateleira do seu chuveiro? Você sabe de todas estas coisas, além de outras milhares, e as pode recordar com grande detalhe. Estas lembranças estão armazenadas em seu córtex. Caberia dizer que estas coisas fazem parte da memória do seu lar. Mas você não pode pensar em todas elas ao mesmo tempo. Sem dúvida, são lembranças relacionadas, mas não há um modo de se lembrar de todos os detalhes de uma só vez. Você tem uma memória completa da sua casa; mas para recordá-la você tem de passar por ela em segmentos consecutivos, de forma muito semelhante a como a experimenta.

Todas as memórias são assim. Tem-se que passar pela sequência temporal das coisas tal e como se fazem. Um padrão (aproximar à porta) recorda o próximo (traspasar a porta), que por sua vez evoca o próximo (dirigir à sala ou subir as escadas), e assim sucessivamente. Cada um é uma sequência do que se seguiu antes. Certamente, fazendo um esforço consciente posso mudar a ordem ao lhe descrever minha casa. Posso saltar do porão ao segundo andar se decido focar-me em artigos sem seguir uma ordem sequencial. Não obstante, uma vez que começo a descrever o cômodo ou o objeto escolhido, volto a repetir uma sequência. Não existem pensamentos realmente aleatórios. A lembrança da memória segue quase sempre uma rota de associação.

Você conhece o alfabeto. Tente o dizer de trás para frente. Você não consegue porque não costuma o experimentar desse modo. Se você deseja saber o que sente uma criança que está aprendendo o alfabeto, tente o dizer de trás para frente. Isso é exatamente o que elas enfrentam, e é duríssimo. Nossa memória do alfabeto é uma sucessão de padrões. Não há nada guardado ou recordado em um instante ou em uma ordem arbitrária. O mesmo acontece com os dias da semana, os meses do ano, os números de telefone e inúmeras outras coisas.

Nossa lembrança das canções constitui um grande exemplo de sequências temporais na memória. Pense em uma melodia que conheça. Gosto de empregar *Somewhere over the Rainbow*, mas qualquer uma servirá. Você não é capaz de imaginar a canção inteira de uma só vez; só em uma sequência. Você pode começar pelo princípio ou talvez com o coro, e depois ir tocando-a, pondo as notas uma a uma. Você não consegue recordar a canção de trás para frente, do mesmo modo que é incapaz de recordá-la do nada. Você a escuta pela primeira vez

enquanto é tocada ao longo do tempo e só é capaz de recordá-la do mesmo modo que a aprendeu.

Isto também é aplicável às lembranças sensoriais de nível sensorial muito baixo. Pensemos sobre nossa memória tátil para as texturas. Nosso córtex possui lembranças da sensação que se produz ao apanhar um punhado de cascalho, deslizar os dedos sobre um tecido de veludo e pressionar uma tecla de piano. Estas memórias baseiam-se em sequências idênticas às do alfabeto e das canções; a única diferença é que estas são mais curtas, duram meras frações de segundo em vez de vários segundos ou minutos. Se enterro sua mão em um balde de cascalho enquanto você dorme, quando acordar não saberá o que estava tocando até que você mova os dedos. Sua memória da textura tátil do cascalho baseia-se em sequências de padrões recolhidos pelos neurônios que percebem a pressão e vibração em sua pele. Estas sequências são diferentes das que receberia se sua mão estivesse enterrada na areia, em bolas de espuma ou em folhas secas. Assim que flexionar a mão, a raspagem e o deslizamento dos cascalhos criariam as sequências de padrões reveladoras de cascalho e desencadeariam a memória apropriada em seu córtex somatosensorial.

Da próxima vez que sair do chuveiro preste atenção ao seu modo de se secar com a toalha. Descobri que eu fazia isto com quase a mesma sucessão exata de esfregues, palmadas e posições corporais todas as vezes. E mediante um experimento aprazível descobri que minha esposa respeita também um padrão semirrígido quando ela sai da ducha. É provável que você também o faça. Se você segue uma sequência, tente mudá-la. Você pode fazer isto, mas precisa se concentrar. Se sua atenção for desviada, voltará a cair em seu padrão de costume.

Todas as memórias se armazenam nas conexões sinápticas que há entre os neurônios. Devido ao grande número de coisas que temos guardado em nosso córtex e que em um determinado momento não podemos recordar mais do que uma fração diminuta das lembranças armazenadas, é lógico pensar que só um número limitado de sinapses e neurônios do nosso cérebro desempenham um papel ativo de cada vez na recuperação da memória. Quando começamos a recordar o que há em nossa casa, é ativado um conjunto de neurônios, que depois faz com que seja ativado outro conjunto de neurônios, e assim sucessivamente. O neocórtex de um humano adulto possui uma capacidade de memória incrivelmente grande. Mas ainda que tenhamos tantas coisas armazenadas, não podemos recordar mais do que algumas ao mesmo tempo, e só seguindo uma sequência de associações.

Realizemos um exercício divertido. Tente recordar detalhes do seu passado, pormenores de onde vivia, lugares que visitou e pessoas que conheceu. Descobri que sempre posso recuperar lembranças de coisas nas quais não tinha pensado durante muitos anos. Há milhares de lembranças detalhadas armazenadas nas sinapses dos nossos cérebros que raramente são usadas. Em um determinado

momento do tempo só recordamos uma fração diminuta do que sabemos. A maior parte da informação espera ociosa os indícios apropriados que a evoquem.

Por regra geral, a memória dos computadores não armazena sequências de padrões. Pode conseguir-se que o faça empregando vários truques de software (como quando se guarda uma canção no computador), mas não é algo automático. Em contraste, o córtex armazena sequências de forma automática. Fazer isto constitui um aspecto inerente do seu sistema de memória neocortical.

* * *

Passemos agora a considerar a segunda característica chave da nossa memória: sua natureza auto-associativa. Como temos visto no capítulo 2, o termo significa simplesmente que os padrões estão associados consigo mesmos. Um sistema de memória auto-associativa é aquele que pode recordar padrões completos quando se lhe dão apenas entradas parciais ou distorcidas. Pode funcionar tanto com padrões espaciais como temporais. Se vemos os pés do nosso filho sobressaindo por trás das cortinas, automaticamente adivinhamos sua forma como um todo. Completamos o padrão espacial com uma versão parcial dele. Ou imaginemos que vemos uma pessoa esperando o ônibus, mas só conseguimos a distinguir em parte porque ela está um pouco tampada por um arbusto. Nosso cérebro não se confunde. Nossos olhos só veem partes de um corpo, mas nosso cérebro preenche o resto, criando uma percepção de uma pessoa completa tão potente que talvez nem sequer nos demos conta de que é uma inferência.

Nós também completamos padrões temporais. Se você recorda um pequeno detalhe de algo que aconteceu há muito tempo, a sequência da lembrança inteira pode chegar a sua mente. A famosa série de novelas de Marcel Proust, *Em busca do tempo perdido*, começava com a lembrança de como ele cheirava uma magdalena e continua se estendendo por mil e uma páginas. Durante a conversa é frequente que não escutemos todas as palavras se nos encontramos em um ambiente barulhento. Não há problema. Nosso cérebro supre o que se nos tem escapado com o que esperamos ouvir. É um fato conhecido que não escutamos todas as palavras que percebemos. Algumas pessoas completam as orações de outros em voz alta, mas em nossas mentes todos nós fazemos isto constantemente. E não só o final das orações, mas também o meio e os inícios. Na maioria das vezes não nos damos conta de que estamos completando padrões de forma contínua, mas é uma característica onipresente e fundamental do modo como se armazenam as lembranças no córtex. Em qualquer momento uma parte pode ativar o todo. Esta é a essência das memórias auto-associativas.

Nosso neocórtex é uma complexa memória auto-associativa biológica. Durante cada momento de estado consciente, cada região funcional está esperando atentamente a chegada de padrões ou fragmentos de padrões conhecidos. Você pode estar concentrado em uma profunda reflexão sobre algo, mas no momento

em que aparece sua amiga, seus pensamentos mudam para ela. Esta mudança não é algo que se decide. O mero aparecimento da sua amiga obriga o seu cérebro a começar a recordar padrões associados com ela. É inevitável. Depois de uma interrupção, é frequente que tenhamos que nos perguntar: "Em que eu estava pensando?". A conversa durante uma comida com amigos segue uma rota tortuosa de associações. A conversa pode começar com os alimentos que temos adiante, mas a salada evoca a lembrança associada daquela salada que fez nossa mãe em nosso casamento, o que leva à lembrança do casamento de outro, que conduz à lembrança de onde foram de viagem de lua de mel, aos problemas políticos dessa parte do mundo, e assim sucessivamente. Os pensamentos e lembranças estão unidos por associação e, mais uma vez, raramente surgem pensamentos aleatórios. As entradas do cérebro unem-se entre si auto-associativamente, completando o presente, e se unem em auto-associação no que costuma continuar a seguir. Chamamos esta corrente de lembranças de *pensamento*, e ainda que seu caminho não seja determinista, também não possuímos um controle pleno a respeito.

* * *

Passemos agora ao terceiro atributo principal da memória neocortical: como ela forma as chamadas representações invariáveis. Neste capítulo me ocuparei das ideias básicas das ditas representações, e no capítulo 6, dos detalhes sobre como o córtex as cria.

A memória de um computador está projetada para armazenar informação da forma exata como ela se apresenta. Se um programa é copiado de um CD para um disco rígido, cada bit é copiado com fidelidade total. Um único erro ou discrepância entre as duas cópias poderia ocasionar a falha do programa. A memória do neocórtex é diferente. Nosso cérebro não recorda com exatidão o que você vê, escuta ou sente. Não recordamos as coisas com fidelidade completa, não porque o córtex e seus neurônios sejam descuidados ou propensos ao erro, mas porque o cérebro recorda as relações importantes no mundo, independentes dos detalhes. Veremos vários exemplos para ilustrar este ponto.

Como temos visto no capítulo 2, durante décadas tem se construído simples modelos de memória auto-associativa e, como acabo de descrever, o cérebro se lembra das lembranças de forma auto-associativa. Mas existe uma grande diferença entre as memórias auto-associativas construídas pelos pesquisadores das redes neurais e as do córtex. As memórias auto-associativas artificiais não empregam representações invariáveis e, portanto, fracassam em alguns aspectos muito básicos. Imaginemos que tenho uma foto de um rosto formado por uma grande acumulação de pontos brancos e negros. Esta foto é um padrão, e se possuo uma memória auto-associativa artificial, posso armazenar muitas fotos de rostos nela. Nossa memória auto-associativa artificial é sólida no sentido de que se lhe proporciono meio rosto ou só um par de olhos ela reconhecerá essa parte da

imagem e completará as partes que faltam corretamente. Este mesmo experimento realizou-se muitas vezes. No entanto, se movo cada ponto da foto cinco pixels à esquerda, a memória falha por completo ao reconhecer o rosto. Para a memória auto-associativa artificial trata-se de um padrão totalmente novo porque nenhum dos pixels entre o padrão guardado com antecedência e o novo estão alinhados. Certamente, nem a você nem a mim nos custaria ver que o padrão mudado é o mesmo rosto. É provável que nem sequer notássemos a mudança. As memórias auto-associativas artificiais não conseguem reconhecer os padrões se forem movidos, rotacionados, mudados de escala ou transformados em algum de outros mil modos, enquanto nosso cérebro maneja essas variações com facilidade. Como podemos perceber que algo é o mesmo ou constante quando os padrões de entrada que o representam são novos e mutáveis? Analisemos outro exemplo.

É provável que neste momento você tenha um livro nas suas mãos. Quando o move, muda a iluminação, se recoloca na cadeira ou fixa os olhos em partes diferentes da página, o padrão de luz que penetra na retina varia por completo. A entrada visual que recebe é diferente um momento após outro e jamais se repete. De fato, poderia segurar este livro durante cem anos e nunca seria exatamente o mesmo o padrão da retina e, portanto, o padrão que entra no cérebro. No entanto, nem por um instante você tem dúvida de que está sustentando um livro, o mesmo livro na realidade. Os padrões internos do seu cérebro que representam “este livro” não mudam, ainda que os estímulos que lhe informam disso estejam em fluxo constante. Por isso que empreguemos o termo *representação invariável* para fazer referência à representação interna do cérebro.

Para citar outro exemplo, pense no rosto de uma amiga. Você o reconhece cada vez que o vê. Acontece de forma automática em menos de um segundo. Não importa se ela se encontra a meio metro, um metro ou do outro lado da sala. Quando ela está perto, sua imagem ocupa a maior parte da sua retina. Quando ela está longe, sua imagem ocupa uma pequena porção desta. Ela pode estar em frente a você, voltada um pouco de lado ou de perfil. Talvez ela esteja sorrindo, entrecerrando os olhos ou bocejando. É provável que você a veja com luz brilhante, em sombra ou sob as luzes fantasmagóricas de ângulos estranhos. Seu semblante pode aparecer em inúmeras posições e variações. Para cada uma o padrão de luz que chega à sua retina é único, mas em todos os casos você sabe instantaneamente que é ela a quem você está olhando.

Abramos o “capô” e olhemos o que acontece dentro do seu cérebro para que ele realize essa façanha espantosa. Sabemos por experimentos que se monitorizamos a atividade dos neurônios da área de entrada visual do seu córtex, chamada V1, o padrão de atividade será diferente para cada visão diferente da face da sua amiga. Cada vez que a face se move ou seus olhos realizam uma nova fixação, o padrão de atividade em V1 muda, do mesmo modo que o faz o da retina. No entanto, se

monitorizamos a atividade das células da sua área de reconhecimento de rostos — uma região funcional que se encontra vários passos acima de V1 na hierarquia cortical—, descobrimos estabilidade. Isto é, alguns dos conjuntos de células da área de reconhecimento visual permanecem ativos enquanto o rosto da sua amiga se encontra dentro do seu campo de visão (ou inclusive enquanto seja evocado por seus olhos mentais), independentemente do seu tamanho, posição, orientação, escala e expressão. Esta estabilidade na atividade celular é uma representação invariável.

Se pensarmos a respeito, esta tarefa parecerá muito simples para merecer ser considerada um problema. É tão automática como respirar. Parece simples porque não somos conscientes do que está acontecendo. E, em certo sentido, é simples porque nossos cérebros podem resolvê-la muito rápido (lembre da regra dos cem passos). No entanto, o problema de compreender como nosso córtex forma representações invariáveis continua sendo um dos maiores mistérios da ciência. É tão difícil?, pergunta-se. Tanto que ninguém, nem sequer usando os computadores mais potentes do mundo, tem sido capaz de resolvê-lo. E não é porque não se tenha tentado.

A especulação sobre este problema possui um histórico antigo. Remonta-se até Platão, há vinte e três séculos. O ateniense perguntava-se como as pessoas eram capazes de pensar e conhecer o mundo. Ele notou que os modelos do mundo real de coisas e ideias sempre são imperfeitos e diferentes. Por exemplo, temos o conceito de um círculo perfeito, mas nunca temos visto um na realidade. Todos os desenhos de círculos são imperfeitos. Inclusive se o desenho com um compasso de geometria, o que chamamos de círculo é representado por uma linha escura, enquanto a circunferência de um círculo de verdade não tem espessura. Como, então, se chega a adquirir o conceito de círculo perfeito? Ou para citar um exemplo mais familiar, pensemos sobre nosso conceito de cães. Qualquer cão que vemos é diferente de todos os demais, e cada vez que vemos o mesmo cão particular obtemos uma visão diferente dele. Todos os cães são diferentes e nunca se pode ver a nenhum cão particular do mesmo modo exato duas vezes. No entanto, todas nossas diversas experiências com cães são canalizadas em um conceito mental de “cão” que é estável para todos eles. O que causava perplexidade a Platão. Como é possível que aprendamos e apliquemos conceitos neste mundo de formas infinitamente variadas e sensações em mudança constante?

A solução do filósofo foi sua famosa Teoria das Formas. Ele chegou à conclusão de que nossas mentes mais elevadas deviam estar amarradas a algum plano transcendente de suprarrealidade, onde existiam ideias fixas e estáveis (Formas, com F maiúsculo) em uma perfeição atemporal. Nossas almas proviam desse lugar místico antes do nascimento, ele decidiu, que é onde elas aprenderam primeiro sobre as Formas. Uma vez que nascemos, conservamos um conhecimento latente delas. A aprendizagem e o entendimento ocorrem porque as formas do mundo real

recordam-nos as Formas com as quais se correspondem. Somos capazes de saber de círculos e cães porque ambos desencadeiam as lembranças das nossas almas dos Círculos e Cães.

Desde uma perspectiva moderna, acaba sendo bastante descabelado. Mas se despojamo-nos da metafísica bombástica, podemos ver que na realidade ele falava da invariância. Seu sistema explicativo estava completamente errado, mas sua intuição de que esta era uma das questões mais importantes que podemos propor sobre nossa natureza foi na mosca.

* * *

Para que não tenham a impressão de que a invariância se reduz à visão, examinemos alguns exemplos de outros sentidos. Pensemos sobre o tato. Quando você enfia a mão no porta-luvas do seu carro para buscar os óculos de sol, seus dedos não têm mais do que esbarrar-se com eles para que você saiba que os encontrou. Não importa que parte da mão estabelece o contato; pode ser o polegar, qualquer parte de um dedo qualquer ou a palma. E o contato pode ser com qualquer parte dos óculos, seja uma lente, a ponte, a haste ou parte do arreiio. Um segundo movimento de qualquer parte da mão sobre qualquer porção dos óculos basta para que o cérebro os identifique. Em cada caso, o fluxo de padrões espaciais e temporais procedente dos seus receptores de tato é completamente diferente —diferentes áreas de pele, diferentes partes do objeto—, mas você apanha os óculos sem pensar nisto.

Ou analisemos a tarefa senso-motora de enfiar a chave no contato do carro. A posição do assento, o corpo, o braço e a mão são levemente diferentes cada vez. A você lhe parece a mesma ação simples e repetitiva dia após dia, mas isto é devido a você ter uma representação invariável dela no cérebro. Se tratassem de fazer um robô que pudesse entrar no carro e enfiar a chave, veriam de imediato o quão difícil que acabaria sendo, a não ser que se assegurem de que o robô estivesse na mesma posição exata e sustente a chave da mesma maneira precisa a cada vez. E inclusive se fosse capaz de conseguir isto, o robô precisaria ser reprogramado para carros diferentes. Aos programas dos robôs e dos computadores, assim como às memórias auto-associativas artificiais, lhes custa muitíssimo manejar a variação.

Outro exemplo interessante é a assinatura. Em algum lugar do córtex motor, no lóbulo frontal, temos uma representação invariável do nosso autógrafo. Cada vez que assinamos nosso nome, usamos a mesma sequência de características, ângulos e ritmos. Fazemos isto minuciosamente com uma caneta esferográfica de ponta fina, de maneira bombástica como John Hancock, no ar com o cotovelo, ou de forma desajeitada com um lápis preso entre os dedos do pé. Parece um pouco diferente a cada vez, certamente, sobretudo nas condições estranhas que acabo de mencionar. No entanto, independentemente da escala, do instrumento de

escritura ou da combinação de partes corporais, sempre empregamos o mesmo “programa motor” abstrato para fazer isto.

Pelo exemplo da assinatura pode-se ver que a representação invariável do córtex motor é, em certos sentidos, a imagem refletida da representação invariável do córtex sensorial. Na parte sensorial, uma ampla variedade de padrões de entrada podem ativar um conjunto estável de células que representa algum padrão abstrato (o rosto da nossa amiga ou nossos óculos de sol). Na parte motora, um conjunto de células estável que representa alguma ordem motora abstrata (apanhar uma bola, assinar nosso nome) é capaz de se expressar utilizando uma ampla variedade de grupos musculares e respeitando uma extensa gama de outras limitações. Esta simetria entre percepção e ação é o que cabe esperar se, como propôs Mountcastle, o córtex executa um único algoritmo básico em todas as áreas.

Voltemos ao córtex sensorial e pensemos novamente sobre a música para analisar um exemplo final. (Gosto de empregar a lembrança da música como exemplo porque é fácil observar todos os temas que o neocórtex deve resolver.) A representação invariável na música ilustra-se por nossa capacidade em reconhecer uma melodia em qualquer tom. O tom em que se toca uma melodia faz referência à escala musical com a qual se compôs. A mesma melodia tocada em diferentes tons começa com notas diferentes. Uma vez que escolhemos o tom para uma interpretação, temos determinado o resto das notas da melodia. Toda melodia pode ser tocada em qualquer tom, o que significa que cada interpretação da “mesma” melodia em um novo tom é na realidade uma sequência de notas completamente diferente. Cada interpretação estimula um conjunto de localizações na cóclea completamente diferente, provocando que um conjunto de padrões espaço-temporais diferentes entrem no córtex auditiva... e, no entanto, percebemos a mesma melodia em cada caso. A não ser que tenhamos um ouvido perfeito, nem sequer seremos capazes de distinguir a mesma canção tocada em dois tons diferentes sem voltar a escutá-las outra vez.

Pensemos na canção *Somewhere over the Rainbow*. É provável que você a aprendeu ao escutar a Judy Garland cantar no filme *O mago de Oz*, mas, a não ser que goze de um ouvido perfeito, não recordará em que tom ela a cantava (lá bemol). Se sento-me no piano e começo a tocar a canção em um tom em que jamais a tenha escutado —digamos em ré—, soará igual. Não se dará conta de que todas as notas são diferentes das da versão que conhecem. Isso significa que sua memória da canção deve estar em uma forma que passa por alto o tom. A memória tem que armazenar as relações importantes da canção, não as notas reais. Neste caso, as relações importantes são o tom relativo das notas, ou “intervalos”. *Somewhere over the Rainbow* começa com uma oitava alta, seguida por um semitono baixo, uma terceira maior baixa, e assim sucessivamente. A estrutura de intervalos da melodia é a mesma para qualquer interpretação, seja

qual for o tom. Sua capacidade para reconhecer a canção em qualquer tom indica que seu cérebro a guardou em sua forma invariável de tom.

De forma similar, a memória do rosto da sua amiga também deve estar armazenada em uma forma que seja independente de qualquer visão particular. O que faz reconhecível seu rosto são suas dimensões relativas, suas cores relativas e suas proporções relativas, não o aspecto que ele apresentou durante um instante no almoço da terça-feira passada. Há "intervalos espaciais" entre as características do seu rosto, do mesmo modo que há "intervalos de tom" entre as notas de uma canção. Seu rosto é largo em relação com seus olhos. Seu nariz é pequeno em relação com a amplitude dos seus olhos. A cor do seu cabelo e a dos seus olhos apresentam uma relação relativa similar que permanece constante ainda que em condições de luz diferentes suas cores absolutas mudem muito. Quando você memorizou o rosto dela, você memorizou esses atributos relativos.

Acho que ocorre uma abstração da forma, similar em todo o córtex, em cada uma das suas regiões. Trata-se de uma propriedade geral do neocórtex. As lembranças armazenam-se em uma forma que capta a essência das relações, não os detalhes do momento. Quando vemos, apalpamos ou escutamos algo, o córtex toma a entrada detalhada e muito específica para converter em uma forma invariável, que é a que se guarda, e é com esta com que se compara cada novo padrão de entrada. O armazenamento da memória, sua recuperação e reconhecimento têm lugar no plano das formas invariáveis. Não existe um conceito equivalente nos computadores.

* * *

Isto chama atenção para um problema interessante. No próximo capítulo sustento que uma função importante do neocórtex é utilizar esta memória para realizar predições. Mas já que o córtex armazena formas invariáveis, como pode efetuar predições específicas? Vejamos alguns exemplos para ilustrar o problema e a solução.

Imagine que seja o ano de 1890 e você está em um povo fronteiriço do Oeste americano. Sua amada está tomando o trem desde o Leste para reunir-se com você em seu novo lar na fronteira. Certamente, você a quer receber na estação quando ela chegar. Durante algumas semanas antes do dia da sua vinda, você se dedica a observar quando chegam os trens e quando saem. Não há horário, e, até onde você tenha podido concluir, o trem nunca entra nem sai na mesma hora durante o dia. Começa a parecer que você não será capaz de predizer quando chegará seu trem. Mas então você se dá conta de que existe certa estrutura nas chegadas e nas saídas dos trens. O trem procedente do Leste chega quatro horas após o que sai nessa direção. Este espaço de quatro horas é constante em um dia após outro, ainda que os tempos específicos variem muito. No dia da sua chegada, você aguarda o aparecimento do trem com destino ao Leste e, quando você o vê,

ajusta o relógio. Quatro horas depois você vai à estação e encontra-se com o trem da sua amada justamente quando ele chega. Esta parábola ilustra tanto o problema que o neocórtex enfrenta como a solução que ele emprega para o resolver.

O mundo tal como o vemos nossos sentidos nunca é o mesmo; assim como a hora de chegada e de saída do trem, sempre é diferente. Compreendemos o mundo buscando uma estrutura invariável no fluxo de entradas em mudança constante. No entanto, esta estrutura invariável não basta para a empregar como base para realizar previsões específicas. Saber que o trem chega quatro horas após o que tem saído não permite você a aparecer na plataforma justamente a tempo para receber a sua amada. Para realizar uma previsão específica, o cérebro deve combinar o conhecimento da estrutura invariável com os detalhes mais recentes. Prever a hora de chegada do trem requer reconhecer a estrutura de quatro horas de intervalo em seu horário e combinar com o conhecimento detalhado da hora em que saiu o último trem com destino ao Leste.

Quando escutamos uma canção conhecida tocada no piano, nosso córtex prediz a próxima nota antes que soe. Mas a memória da canção, como temos visto, está em uma forma invariável de tom. Nossa memória indica-nos qual é o próximo intervalo, mas não nos diz nada sobre a nota real. Para prever a nota exata que se seguirá é preciso combinar o próximo intervalo com a última nota específica. Se o próximo intervalo é uma terceira maior e a última nota que temos escutado foi dó, cabe prever que a próxima nota específica será mi. Escutamos em nossa mente mi, não "terceira maior". E a não ser que tenhamos identificado mau a canção ou o pianista tenha se equivocado, nossa previsão está correta.

Ao ver o rosto da nossa amiga, nosso córtex contribui e prediz a miríade de detalhes da sua imagem única nesse instante. Comprova que seus olhos são perfeitos, e que seu nariz, lábios e cabelo são como devem ser. Nosso córtex efetua estas previsões com uma grande especificidade. É capaz de prever detalhes insignificantes sobre seu rosto ainda que jamais o tenhamos visto antes nessa orientação ou ambiente particular. Se temos a noção de onde estão os olhos e nariz da nossa amiga, e conhecemos a estrutura do seu rosto, podemos prever com exatidão onde devem estar seus lábios. Se sabemos que sua pele está tingida de laranja pela luz do pôr-do-sol, sabemos em que cor deve aparecer seu cabelo. Mais uma vez, nosso cérebro faz isto combinando a memória da estrutura invariável do seu rosto com os detalhes particulares da nossa experiência imediata.

O exemplo do horário do trem não é mais do que uma analogia do que acontece em nosso córtex, mas os exemplos da melodia e do rosto, não. A combinação de representações invariáveis e de entradas imediatas para estabelecer previsões detalhadas é exatamente o que acontece. É um processo onipresente que ocorre em todas as regiões do córtex. É assim que efetuamos previsões específicas sobre o cômodo em que estamos sentados neste momento. É assim que somos capazes

de prever não só as palavras que dirão os demais, mas também o tom de voz que empregarão para o fazer, seu sotaque e de que parte do cômodo esperamos os escutar. É assim que sabemos com precisão quando nosso pé dará no solo e como será subir um trecho de degraus. É assim que somos capazes de assinar nosso nome com o pé, ou apanhar uma bola que nos lançam.

As três propriedades da memória cortical analisadas neste capítulo (armazenamento de sequências, memória auto-associativa e representações invariáveis) são ingredientes necessários para prever o futuro baseado em lembranças do passado. No próximo capítulo proponho que efetuar previsões constitui a essência da inteligência.

Um Novo Modelo para a Inteligência

Em um dia de abril de 1986 pus-me a pensar sobre o que significava “entender” algo. Tinha-me passado meses lutando com a pergunta fundamental do que fazem os cérebros quando não geram comportamento. O que faz um cérebro quando escuta de forma passiva um discurso? O que faz agora mesmo seu cérebro enquanto você está lendo? A informação entra no cérebro, mas não sai. O que lhe acontece? Seus comportamentos neste momento são provavelmente básicos — como a respiração e os movimentos oculares—, mas quando você se encontra em estado consciente seu cérebro faz bem mais do que isso enquanto você lê e compreende estas palavras. Compreender deve ser o resultado da atividade neural. Mas qual? O que fazem os neurônios quando entendem?

Ao olhar ao redor em meu escritório nesse dia vi objetos conhecidos: cadeiras, cartazes, janelas, plantas, lápis e assim por diante. Rodeavam-me centenas de artigos. Meus olhos viam-nos enquanto olhava em torno, mas só o fato de vê-los não me incitava a realizar uma ação. Não se invocava ou requeria nenhum comportamento, mas de certo modo eu “entendia” a sala e seu conteúdo. Eu estava fazendo o que não podia fazer a sala chinesa de Searle, e sem ter que passar nada através de uma fenda. Compreendia, mas não tinha nenhuma ação que demonstrasse isto. O que significava “compreender”?

Foi enquanto meditava sobre este dilema que tive uma percepção, um desses intensos momentos emocionais em que de repente o que era um emaranhado de confusão se torna claro e compreensível. Tudo o que fiz foi me propor sobre o que aconteceria se um novo objeto, algo que eu nunca tivesse visto antes, aparecesse na sala, digamos uma caneca de café azul.

A resposta parecia simples. Me daria conta de que o novo objeto não correspondia a esse lugar. Chamaria minha atenção por ser novo. Não precisaria me perguntar de forma consciente se a caneca de café era nova. Se limitaria a ressaltar como algo que pertencia ao lugar. Subjacente a esta resposta aparentemente trivial está um conceito poderoso. Para apreciar que algo é diferente, certos neurônios do meu cérebro antes inativos teriam entrado em funcionamento. Como saberiam os ditos neurônios que a caneca de café azul era nova e as centenas de objetos restantes que tinha na sala não? A resposta continua surpreendendo-me. Nosso cérebro emprega memórias armazenadas para realizar predições constantes sobretudo no que vemos, sentimos e escutamos. Quando olho ao redor da sala, meu cérebro está utilizando lembranças para formar predições sobre o que espera experimentar antes que aconteça. A vasta maioria das predições ocorrem sem que tenhamos consciência disso. É como se diferentes partes do meu cérebro estivessem dizendo: “Está o computador no meio do escritório? Sim. É preto? Sim. Está o lustre no canto direito do escritório? Sim. Está o dicionário onde o deixei?

Sim. É a janela retangular e a parede vertical? Sim. Entra a luz do sol desde a direção adequada para a hora do dia? Sim.” Mas quando entra um padrão visual que não tinha memorizado nesse contexto, a predição não se cumpre e minha atenção se dirige ao erro.

Certamente, o cérebro não fala consigo mesmo enquanto realiza predições, e não as faz em série. Também não limita-se a realizar predições sobre diferentes objetos como canecas de café. Nosso cérebro efetua predições constantes sobre a mesma estrutura do mundo em que vivemos, e o faz em paralelo. Ele facilmente detectará uma textura estranha, um nariz deformado ou um movimento inusual. O caráter omnipresente destas predições, inconscientes em sua maioria, não se adverte a primeira vista, motivo pelo qual talvez temos passado por alto sua importância durante tanto tempo. Acontecem de modo tão automático, com tanta facilidade, que não conseguimos desvendar o que está acontecendo dentro dos nossos crânios. Espero conseguir lhe comunicar a força desta ideia. A predição é tão dominante que o que “percebemos” —isto é, como aparece diante de nós o mundo— não provém unicamente dos nossos sentidos. O que percebemos é uma combinação do que apreciamos e das predições do nosso cérebro derivadas da memória.

* * *

Minutos depois concebi um experimento mental para ajudar a transmitir o que tinha compreendido nesse momento. Chamo-o de experimento da porta modificada. É como segue.

Quando você chega em casa todo dia, costuma demorar alguns segundos em traspasar a porta principal ou qualquer outra porta que use. Você chega até ela, gira a maçaneta, entra, e a fecha por trás de você. É um hábito firmemente estabelecido, algo que você faz de forma constante e que presta escassa atenção. Suponhamos que enquanto você se encontra fora, eu me introduza em sua casa e mude algo da sua porta. Poderia ser qualquer coisa. Poderia deslocar a maçaneta um milímetro, trocar uma maçaneta redonda por outra alongada, ou uma de latão por outra cromada. Poderia variar o peso da porta, substituindo o carvalho maciço por laminado, ou vice-versa. Poderia fazer com que as dobradiças chiassem ou que se deslizassem sem fricção. Poderia ampliar ou reduzir a largura da porta e seu modelo. Poderia mudar sua cor, acrescentar uma aldrava onde costumava estar o olho mágico, ou uma janela. Sou capaz de imaginar milhares de mudanças que fariam com que sua porta fosse desconhecida para você. Quando você chegasse a sua casa nesse dia e tentasse a abrir, detectaria em seguida que algo vai mau. Talvez lhe levaria uns segundos de reflexão para se dar conta exatamente do que seja, mas perceberia a mudança com muita rapidez. Quando sua mão atingisse a maçaneta deslocada, notaria que ela não está na localização correta. Ou quando visse a nova janela da porta, algo lhe pareceria raro. Ou se foi mudado seu peso,

não empurraria com a força adequada e se surpreenderia. O caso é que você se dará conta de um dos milhares de mudanças em um lapso de tempo muito curto.

Como você faz isto? Como você percebe essas mudanças? Os engenheiros da computação ou da inteligência artificial abordariam este problema criando uma lista de todas as propriedades da porta e as pondo em uma base de dados, com campos para cada atributo que uma porta pode ter e entradas específicas para sua porta particular. Quando você se aproximasse da porta, o computador verificaria em toda a base de dados, buscando largura, cor, tamanho, posição da maçaneta, peso, som e assim por diante. Ainda que superficialmente esta operação possa parecer similar ao modo como tenho descrito de que meu cérebro comprovava cada uma da sua miríade de predições enquanto olhava ao redor em meu escritório, a diferença é real e de longo alcance. A estratégia da inteligência artificial não é plausível. Em primeiro lugar, é impossível especificar antecipadamente todos os atributos que uma porta pode ter. A lista é em potencial interminável. Em segundo lugar, precisaríamos contar com listas similares para cada objeto que nós encontrássemos a cada segundo das nossas vidas. Em terceiro lugar, nada do que sabemos sobre cérebros e neurônios sugere que seja bem assim como funciona. E, por último, os neurônios são muito lentos para aplicar bases de dados do tipo que empregam os computadores. Você demoraria vinte minutos em vez de dois segundos para dar-se conta das mudanças quando você passasse pela porta.

Só há um modo de interpretar sua reação diante da porta modificada: seu cérebro realiza predições sensoriais de baixo nível sobre o que espera ver, escutar e sentir em cada determinado momento, e faz isto em paralelo. Todas as regiões do seu neocórtex tentam prever ao mesmo tempo qual será sua próxima experiência. As áreas visuais efetuam predições sobre bordas, formas, objetos, posições e movimentos; as áreas auditivas, sobre tons, direção da fonte e padrões de som; e as áreas somatosensoriais, sobre tato, textura, contorno e temperatura.

A "predição" significa que os neurônios que participam na apreciação da porta se ativam antes de receber a entrada sensorial; quando chega esta, é comparada com o que se esperava. Enquanto você se aproxima da porta, seu córtex vai formando um montão de predições baseadas na experiência passada. Quando você chega a ela, ele prediz o que sentirá nos dedos quando a tocar e em que ângulo estarão suas articulações quando fizer isto. Quando começa a empurrar para a abrir, seu córtex prediz quanta resistência oferecerá e como soará. Quando se cumprem todas suas predições, você a atravessará sem ter consciência de que as ditas predições se verificaram. Mas se suas expectativas são violadas, o erro fará com que se dê conta. As predições corretas dão como resultado o entendimento. A porta está normal. As predições errôneas induzem a confusão e suscitam sua atenção. A maçaneta da porta não está onde deveria. A porta está muito leve. A porta está descentralizada. A textura da maçaneta não é a

adequada. Realizamos previsões contínuas de baixo nível em paralelo com todos nossos sentidos.

Mas isso não é tudo. Minha proposta é bem mais ambiciosa. A previsão não é só uma das coisas que faz nosso cérebro. É a função primordial do neocórtex e a base da inteligência. O córtex é um órgão de previsão. Se queremos entender o que é a inteligência, o que é a criatividade, como funciona nosso cérebro e como construir máquinas inteligentes, devemos compreender a natureza destas previsões e como as realiza o córtex. Até o comportamento entende-se melhor como um produto derivado da previsão.

* * *

Não sei quem foi a primeira pessoa que sugeriu que a previsão é a chave para compreender a inteligência. Na ciência e na indústria, ninguém inventa nada completamente novo; pelo contrário, trata-se bem mais de amoldar ideias existentes a novos modelos. Os componentes de uma nova ideia costumam estar flutuando no meio do discurso científico antes da sua descoberta; o que costuma ser novo é a coesão desses componentes em um conjunto abrangedor. Da mesma forma, a ideia de que a função primordial do córtex é realizar previsões não é completamente nova. Tem estado flutuando ao redor em várias formas durante algum tempo, mas ainda não tem assumido sua posição legítima no centro da teoria do cérebro e da definição da inteligência.

Acaba sendo irônico que alguns dos pioneiros da inteligência artificial tivessem a noção de que os computadores deviam construir um modelo do mundo e o utilizar para realizar previsões. Em 1956, por exemplo, D. M. Mackay sustentou que as máquinas inteligentes deveriam ter um “mecanismo de resposta interno” projetado para “combinar o que se recebe”. Ele não empregou as palavras “memória” e “previsão”, mas seu pensamento ia nessa linha.

Desde meados da década de 1990, termos como *inferência*, *modelos generativos* e *previsão* têm entrado no vocabulário científico. Todos se referem a ideias relacionadas. Como exemplo, em seu livro de 2001, *O cérebro e o mito do eu*, Rodolfo Llinás, da Escola de Medicina da Universidade de Nova York, escreveu: “A capacidade para prever o resultado de acontecimentos futuros —crucial para que o movimento tenha sucesso— é muito provável que seja a função primordial e mais comum de todas as funções globais do cérebro”. Cientistas como David Mumford da Universidade Brown, Rajesh Rao da Universidade de Washington, Stephen Grossberg da Universidade de Boston e muitos outros têm escrito e teorizado sobre o papel da realimentação e da previsão de vários modos. Há um subcampo inteiro na matemática dedicado às redes bayesianas. Batizadas desse modo por Thomas Bayes, pastor inglês nascido em 1702 que foi pioneiro em estatística, as redes bayesianas utilizam a teoria da probabilidade para efetuar previsões.

O que tem faltado é colocar estes elementos distintos dentro de um modelo teórico coerente. Sustento que isto não foi feito antes, e constitui a meta deste livro.

* * *

Antes de entrar em detalhes sobre como realiza previsões o córtex, analisemos alguns exemplos adicionais. Quanto mais pensarem nesta ideia, mais se darão conta de que a previsão é omnipresente e a base do nosso modo de compreender o mundo.

Esta manhã fiz panquecas. Em um ponto do processo, alonguei a mão por debaixo da mesa para abrir a porta de um armário. Por intuição sabia sem vê-lo qual seria o tato —neste caso, o da maçaneta do armário— e quando o sentiria. Torci a tampa da embalagem do leite esperando que ele virasse e então caísse livremente. Acendi a chapa esperando pressionar o botão levemente e logo depois girá-lo com certa resistência. Esperava escutar o suave som do lume de gás mais ou menos um segundo depois. Em cada minuto passado na cozinha realizei dezenas ou centenas de movimentos, e cada um representava muitas previsões. Sei isto porque se algum destes movimentos comuns tivesse tido um resultado diferente do esperado me teria dado conta.

Cada vez que você põe o pé no solo enquanto caminha, seu cérebro prediz quando o mesmo se deterá e quanta “elasticidade” terá o material em que pisar. Se alguma vez falhou-lhe um degrau em um trecho de escadas, sabe com quanta rapidez se dá conta de que algo vai mau. Você baixa o pé e no momento em que “passa através” do degrau previsto você sabe que tem dificuldades. O pé não aprecia nada, mas seu cérebro fez uma previsão que não se cumpriu. Um robô dirigido por computador cairia facilmente ao não se dar conta de que acontecia algo fora da ordem, enquanto você o saberia tão cedo caso seu pé ultrapassasse em uma fração de milímetro o lugar onde seu cérebro tem esperado que você se detenha.

Quando você escuta uma melodia conhecida, ouve a próxima nota em sua cabeça antes que ela soe. Quando você escuta um álbum que gosta, você ouve do começo de cada canção um par de segundos antes que se inicie. O que acontece? Os neurônios do seu cérebro que se estimulariam quando você escutasse a próxima nota o fazem por adiantado e, deste modo, você “escuta” a canção em sua cabeça. Os neurônios estimulam-se em resposta à memória, que pode ser muito duradoura. Não é inusual escutar um álbum de música pela primeira vez em muitos anos e continuar ouvindo a próxima canção de forma automática antes que a anterior tenha acabado. E cria-se uma agradável sensação de leve incerteza quando se escuta o álbum favorito soando a esmo, pois se sabe que a previsão da próxima canção errará.

Quando escutamos alguém falar, com frequência sabemos o que ele vai dizer antes que tenha terminado, ou pelo menos pensamos que o sabemos. Às vezes nem sequer escutamos o que diz o orador na realidade, mas o que esperamos escutar. (Isto me acontecia com tanta frequência quando era pequeno que minha mãe me levou ao médico para que fosse revisada a minha audição.) E é assim porque a gente tende a empregar expressões comuns em boa parte da sua conversa. Se digo: "*How now brown...*", seu cérebro ativará os neurônios que representam a palavra *cow* antes que a pronuncie (se o inglês não é sua língua nativa, talvez não tenha ideia do que estou falando. Trata-se de um conhecido jogo de palavras que significa o que acontece agora). Certamente, não sabemos todas as vezes o que vão dizer os demais. A predição nem sempre é exata, pois nossas mentes funcionam realizando predições prováveis sobre o que está a ponto de acontecer. Às vezes conhecemos com exatidão o que vai ocorrer, outras vezes nossas expectativas se distribuem entre várias possibilidades. Se estivéssemos falando em uma mesa durante uma comida e eu dissesse: "Por gentileza, passa-me a...", seu cérebro não se surpreenderia se acrescentasse a seguir "sal", "pimenta" ou "mostarda". Em certo sentido, seu cérebro prediz todos estes resultados possíveis ao mesmo tempo. No entanto, se eu dissesse: "Por gentileza, passa-me a calçada", você saberia que algo vai mau.

Voltando à música, nela também podemos ver predições prováveis. Se estamos escutando uma canção que jamais a tenhamos ouvido antes, de todos modos podemos ter previsões bastante sólidas. Na música *country* espero um compasso regular, um ritmo repetido, frases que duram o mesmo número de compassos e canções que terminem em tom de altura. Talvez você não saiba o que significam estes termos, mas —supondo que tenha escutado uma música similar— seu cérebro prediz de forma automática compassos, ritmos repetidos, terminação das frases e finais das canções. Se uma nova canção viola estes princípios, você sabe de imediato que algo vai mau. Pense nisso durante um segundo. Escute uma canção que jamais tenha ouvido, seu cérebro experimenta um padrão que jamais tenha experimentado, e no entanto realiza predições e pode dizer se algo vai mau. A base destas predições em boa parte inconscientes é o conjunto de memórias que estão armazenadas em nosso córtex. Nosso cérebro não pode afirmar com exatidão o que acontecerá a seguir, mas prediz qual padrão de nota é provável que toque e qual não.

Todos temos tido a experiência de nos dar conta de improviso de que uma fonte de ruído de fundo constante, como uma britadeira ou uma escavadeira, acaba de cessar; no entanto, não tínhamos percebido o som enquanto ocorria. Nossas áreas auditivas estavam predizendo sua continuação um momento após outro, e enquanto o som não mudou, não lhe prestamos atenção. Ao cessar, nossa predição foi violada e isto atraiu nossa atenção. Vejamos um exemplo histórico. Logo após que os trens elevados na cidade de Nova York deixassem de funcionar, as pessoas chamavam à polícia no meio da noite para declararem que algo as

tinham acordado. Costumavam efetuar a chamada aproximadamente à mesma hora em que os trens costumavam passar por seus apartamentos.

Gostamos de afirmar que ver é crer. Não obstante, vemos o que esperamos ver com tanta frequência como vemos o que na realidade vemos. Um dos exemplos mais fascinantes a respeito tem que ver com o que os pesquisadores chamam de completar. É provável que você já esteja inteirado de que temos um pequeno ponto cego em cada olho, onde o nervo óptico sai de cada retina por um buraco chamado papila óptica. Nessa zona não contamos com fotorreceptores, de modo que estamos permanentemente cegos no ponto correspondente do campo visual. Existem duas razões pelas quais não costumamos nos dar conta disso, uma trivial e a outra instrutiva. A trivial é que nossos dois pontos cegos não costumam se sobrepor, assim um olho compensa o outro.

Mas acaba sendo interessante que também não percebamos o ponto cego quando só temos aberto um olho. Nosso sistema visual "completa" a informação que falta. Quando fechamos um olho e olhamos um luxuoso tapete turco ou os contornos ondulantes da textura da madeira em uma mesa de cerejeira, não vemos um buraco. Nós escuros inteiros no tapete, nós escuros inteiros na textura da madeira desaparecem da visão da nossa retina quando dá a casualidade de coincidirem com os pontos cegos, mas nossa experiência é uma extensão sem costuras de texturas e cores. Nosso córtex visual recorre às lembranças de padrões similares e efetua uma corrente constante de predições que completam qualquer entrada que falte.

Completa-se em todas partes da imagem visual, não só nos pontos cegos. Por exemplo, mostro-lhe uma foto de uma praia com um tronco arrastado pela corrente sobre umas rochas. O limite entre as rochas e o tronco parece claro e evidente. No entanto, se ampliamos a imagem, veremos que as rochas e o tronco apresentam uma textura e cor similares onde se encontram. Na foto ampliada, a borda do tronco não parece distinguível das rochas. Se observamos a cena por inteiro, a borda do tronco está clara, mas na realidade a inferimos do resto da imagem. Quando olhamos o mundo, percebemos linhas e limites claros que separam os objetos, mas os dados brutos que entram em nossos olhos são com frequência ruidosos e ambíguos. Nosso córtex completa as seções que faltam ou estão desordenadas com o que ele pensa que deve ser. E percebemos uma imagem inequívoca.

A predição na visão é também uma função do modo de se mover nossos olhos. No capítulo 3 tenho mencionado as sacadas oculares. Uma três vezes a cada segundo, nossos olhos fixam-se em um ponto e depois saltam de repente a outro. Em geral, não nos damos conta destes movimentos e normalmente não os controlamos de forma consciente. E cada vez que nossos olhos se fixam em um novo ponto, o padrão que entra em nosso cérebro muda por completo desde a última fixação. Portanto, três vezes por segundo nosso cérebro vê algo muito

diferente. As sacadas oculares não são completamente aleatórias. Quando olhamos um rosto, nossos olhos costumam se fixar primeiro em um olho, e depois no outro; vão de um lado a outro e se fixam de forma ocasional no nariz, na boca, nas orelhas e em outras características. Só percebemos o "rosto", mas os olhos veem um olho, outro olho, o nariz, a boca, um olho, e assim sucessivamente. Dou-me conta de que lhe parece que não é assim. O que você percebe é uma visão contínua do mundo, mas os dados brutos que entram na sua cabeça estão tão entrecortados como se tivessem sido gravados com uma câmera de vídeo mau manejada.

Imaginemos agora que nós encontramos alguém com um nariz adicional no lugar onde devia ter um olho. Nossos olhos fixam-se primeiro em um olho, e depois de uma sacada ocular, no outro, mas no seu lugar vemos um nariz. Saberíamos sem dúvida que algo ia mau. Para que isto aconteça, nosso cérebro deve ter uma expectativa ou predição do que está a ponto de ver. Quando predizemos olho, mas vemos nariz, não se cumpre a predição. Portanto, várias vezes por segundo, coincidindo com cada sacada ocular, nosso cérebro realiza uma predição sobre o que verá a seguir. Quando esta predição é errônea, se suscita de imediato nossa atenção. Por isso nos parece difícil não olhar às pessoas com deformidades. Se você visse uma pessoa com dois narizes, lhe custaria não ficar a olhando fixamente? Certamente, se vivêssemos com a dita pessoa, depois de um período nos acabaríamos acostumando com os dois narizes e nos deixaria de parecer incomum.

Pense agora em você mesmo. Que predições está realizando? Quando você passa as páginas deste livro, tem expectativas de que elas se dobrarão um pouco e se moverão de formas predizíveis diferentes do modo de se mover a capa. Se você está sentado, está predizendo que a sensação de pressão em seu corpo persistirá; mas se o assento se umedecesse, começasse a deslizar-se para atrás ou sofresse qualquer outra mudança inesperada, você deixaria de prestar atenção no livro e tentaria imaginar o que está acontecendo. Se você emprega certo tempo em observar-se, pode começar a entender que sua percepção do mundo, seu entendimento do mundo, está intimamente unida à predição. Seu cérebro tem feito um modelo do mundo e está o combinando de forma constante com a realidade. Você sabe onde se encontra e o que está fazendo pela validade desse modelo.

A predição não se limita a padrões de informação sensorial de baixo nível como ver e escutar. Até agora tenho restringido a discussão a tais exemplos porque são o modo mais simples de apresentar este modelo para compreender a inteligência. No entanto, segundo o princípio de Mountcastle, o que é válido para as áreas sensoriais de nível baixo também deve ser para todas as áreas corticais. O cérebro humano é mais inteligente que o de outros animais porque pode realizar predições sobre tipos de padrões mais abstratos e sequências de padrões temporais mais

longas. Para predizer o que dirá minha esposa quando ela me ver, devo saber o que ela tem dito no passado, que hoje é sexta-feira, que a lixeira tem que ser posta no meio-fio nas sextas-feiras pela noite, que não o fiz a tempo na semana passada e que seu rosto tem certa aparência. Quando ela abre a boca, tenho uma predição bastante sólida do que ela vai dizer. Neste caso, não sei quais serão as palavras exatas, mas sei que ela me recordará de retirar a lixeira. O aspecto importante é que a inteligência superior não é um tipo diferente de processo da inteligência perceptiva. Baseia-se fundamentalmente na mesma memória neocortical e algoritmo de predição.

Temos de advertir que nossos testes de inteligência são em essência provas de predição. Do jardim de infância à faculdade, as provas de coeficiente intelectual baseiam-se em realizar predições. Dada uma sequência de números, qual deve ser a próxima? Dadas três visões diferentes de um objeto com uma forma complexa, qual das seguintes corresponde também ao objeto? A palavra A é à palavra B o que a palavra C é a que palavra?

A ciência é em si mesma um exercício de predição. Adiantamos nosso conhecimento do mundo mediante um processo de hipótese e verificação. Este livro é em essência uma predição sobre o que é a inteligência e como funciona o cérebro. Inclusive o projeto de produtos é sobretudo um processo preditivo. Ao projetar roupa ou automóveis, os engenheiros tentam predizer o que farão os concorrentes, o que desejarão os consumidores, quanto custará um novo projeto e que modas terão demanda.

A inteligência mede-se pela capacidade de recordar e predizer padrões do mundo, incluídos linguagem, matemática, propriedades físicas dos objetos e situações sociais. Nosso cérebro percebe padrões do mundo exterior, armazena-os como memória e realiza predições baseadas nas comparações entre o que tem visto antes e o que acontece agora.

* * *

Neste ponto você talvez esteja pensando: "Aceito que posso ser inteligente simplesmente deitado na escuridão. Como tem assinalado, não preciso agir para compreender ou ser inteligente. Mas não são essas situações a exceção? Sustenta realmente que o entendimento e o comportamento inteligentes estão completamente separados? No final, não é o comportamento, e não a predição, o que nos faz inteligentes? Além disso, o comportamento é o determinante supremo da sobrevivência".

É uma pergunta razoável e, certamente, no final o comportamento é o que mais importa para a sobrevivência de um animal. Predição e comportamento não estão separados por completo, mas sua relação é sutil. Em primeiro lugar, o neocórtex apareceu no cenário evolutivo quando os animais já tinham desenvolvido

comportamentos sofisticados. Portanto, seu valor de sobrevivência deve entender-se antes de mais nada atendendo às diversas melhoras que podia conferir aos comportamentos animais existentes. Primeiro chegou o comportamento; depois, a inteligência. Em segundo lugar, menos no caso da audição, a maioria do que sentimos depende em boa medida do nosso modo de se mover no mundo. Portanto, a predição e o comportamento estão estreitamente relacionados. Exploremos estes temas.

Os mamíferos desenvolveram um grande neocórtex porque isto proporcionava-lhes certa vantagem de sobrevivência, e esta vantagem acabou se implantando no comportamento. Mas no começo o córtex servia para fazer um uso mais efetivo dos comportamentos existentes, não para criar comportamentos completamente novos. Para tornar claro este ponto é preciso observar como evoluíram nossos cérebros.

Surgiram sistemas nervosos simples não muito após as criaturas multicelulares começarem a pulular por toda a Terra, há centenas de milhões de anos; mas a história da inteligência real começa em data mais recente com nossos antepassados répteis, que conseguiram conquistar a Terra. Eles espalharam-se por todos os continentes e se diversificaram em numerosas espécies. Possuíam sentidos agudos e cérebros bem desenvolvidos que lhes dotavam de um comportamento complexo. Seus descendentes diretos, os répteis que sobrevivem na atualidade, continuam os conservando. Um jacaré, por exemplo, goza de sentidos sofisticados como os seus ou os meus. Possui olhos, ouvidos, nariz, boca e pele bem desenvolvidos. Realiza comportamentos complexos, entre os quais se incluem sua capacidade para nadar, correr, se ocultar, caçar, armar emboscadas, tomar sol, aninhar e nivelar-se.

Que diferença há entre um cérebro humano e um de réptil? Muita e pouca. Digo pouca porque, em linhas gerais, tudo o que há em um cérebro de réptil existe em um humano. Digo muita porque um cérebro humano possui algo importantíssimo que um réptil não tem: um grande córtex. Às vezes escutamos as pessoas se referirem ao cérebro "velho" ou cérebro "primitivo". Todo ser humano conta com estas estruturas mais antigas no cérebro, igual a um réptil. Elas regulam a pressão sanguínea, a fome, o sexo, as emoções e muitos aspectos do movimento. Quando estamos de pé, nos balançamos e andamos, por exemplo, dependemos muito do cérebro velho. Se escutamos um som aterrador, temos medo e começamos a correr, e isto se deve em boa medida ao nosso cérebro velho. Não se precisa mais do que um cérebro de réptil para fazer muitas coisas interessantes e úteis. Portanto, o que faz o neocórtex se ele não é requerido estritamente para ver, escutar e se mover?

Os mamíferos são mais inteligentes que os répteis devido ao seu neocórtex, que apareceu há dezenas de milhões de anos e é exclusivo deles. O que faz os humanos mais inteligentes do que o restante dos mamíferos é sobretudo a grande

zona do neocórtex, que se estendeu de forma espetacular há só um par de milhões de anos. Recordemos que a constituição do córtex emprega um único elemento repetido. A lâmina do córtex humano é da mesma espessura e tem praticamente a mesma estrutura que a dos nossos parentes mamíferos. Quando a evolução faz algo grande com muita rapidez, como no caso do córtex humano, o consegue copiando uma estrutura existente. Tornamo-nos inteligentes acrescentando muitos mais elementos a um algoritmo cortical comum. Um erro habitual consiste em considerar que o cérebro humano é o ponto mais alto de bilhões de anos de evolução, o que pode ser verdade se pensamos no sistema nervoso como um todo. No entanto, o neocórtex é uma estrutura relativamente nova e não leva em vigor o tempo suficiente para ter passado pelo refinamento evolutivo de longo prazo.

Este é o núcleo do meu argumento sobre como entender o neocórtex e por que a memória e a predição são as chaves para revelar o mistério da inteligência. Começamos com o cérebro dos répteis sem córtex. A evolução descobre que se for agregado um sistema de memória (o neocórtex) à rota sensorial do cérebro primitivo o animal adquire a capacidade de predizer o futuro. Imaginemos que o antigo cérebro de réptil continua fazendo suas coisas, mas agora ao mesmo tempo se introduzem padrões sensoriais no neocórtex, que guarda esta informação em sua memória. Em um tempo futuro, quando o animal se encontra na mesma situação ou em outra similar, a memória reconhece as entradas e recorda o que aconteceu no passado. A memória recordada é comparada com o fluxo de entradas sensoriais; “completa” a entrada presente e prediz o que se verá a seguir. Ao comparar a entrada sensorial presente com a memória recordada, o animal não só compreende onde está, mas que pode ver o futuro.

Agora imaginemos que o córtex recorda não só o que o animal tem visto, mas também os comportamentos que o antigo cérebro realizava quando se achava em uma situação similar. Nem sequer temos de supor que o córtex conhece a diferença entre sensações e comportamento; para o córtex ambos não são mais do que padrões. Quando nosso animal se encontra na mesma situação ou em outra parecida, não só vê o futuro, mas que recorda que comportamentos conduziram a essa visão do futuro. Deste modo, a memória e a predição permitem a um animal utilizar seus comportamentos existentes (o cérebro velho) de forma mais inteligente.

Por exemplo, imagine que você seja um rato que está aprendendo a se orientar por um labirinto pela primeira vez. Excitado pela incerteza ou pela fome, você utilizará as faculdades inerentes do seu cérebro velho para explorar o novo ambiente: escutar, olhar, cheirar e rastejar próximo aos muros. Toda esta informação sensorial é utilizada por seu cérebro velho, mas também passa ao neocórtex, onde se armazena. Em algum momento futuro você se encontrará no mesmo labirinto. Seu neocórtex reconhecerá a entrada presente como algo que já

tinha visto e recordará os padrões guardados que representam o que aconteceu no passado. Em essência, permite-lhe ver um trecho curto do futuro. Se você fosse um rato que fala, diria: "Oh, reconheço este labirinto, e me lembro dessa esquina". Quando seu neocórtex se lembrar do que aconteceu no passado, se imaginará encontrando o queijo que viu pela última vez no labirinto e como chegou até ele. "Se giro aqui, sei o que acontecerá depois. Há um pedaço de queijo no final deste corredor. Vejo-o em meu imaginação." Enquanto corre pelo labirinto, você depende de estruturas mais antigas e primitivas para efetuar movimentos como levantar as patas e limpar os bigodes. Com seu neocórtex (relativamente) grande você é capaz de recordar os lugares nos quais tenha estado, os voltar a reconhecer no futuro e efetuar predições sobre o que vai acontecer a seguir. Uma lagartixa sem neocórtex possui uma capacidade muito menor para recordar o passado e talvez tenha que se orientar novamente a cada vez no labirinto. Você (o rato) compreende o mundo e o futuro imediato devido a sua memória cortical. Vê imagens gráficas das recompensas e perigos que há para cada decisão e, portanto, se move com maior efetividade por seu mundo. Você pode ver, literalmente, o futuro.

Mas tenha em conta que você não está realizando comportamentos que sejam muito complexos nem novos. Você não está construindo uma asa delta e voando para o queijo do final do corredor. Seu neocórtex está formando predições sobre os padrões sensoriais que lhe permitem ver o futuro, mas seu leque de comportamentos disponíveis não tem sofrido variações. Sua faculdade para correr, trepar e explorar continua parecendo-se muito à de uma lagartixa.

Quando o córtex foi se alargando ao longo do período evolutivo, ele pôde ser capaz de recordar cada vez mais sobre o mundo. Pôde formar mais memórias e realizar mais predições. A complexidade das ditas memórias e predições também aumentou. Mas aconteceu algo mais notável que conduziu à capacidade única dos seres humanos de ter um comportamento inteligente.

O comportamento humano evidencia o antigo repertório básico de deslocar-se ao redor com habilidades de rato. Temos levado a evolução neocortical a um plano novo. Só os seres humanos criam linguagem escrita e falada. Só os humanos cozinham seus alimentos, fazem suas roupas, voam em aviões e constroem arranha-céus. Nossas faculdades motoras e planejadoras excedem significativamente as dos nossos parentes animais mais próximos. Como pode o córtex, que foi projetado para realizar predições sensoriais, gerar o comportamento incrivelmente complexo único dos humanos? E como pôde evoluir tão de repente este comportamento superior? Há duas respostas para estas perguntas. Uma é que o algoritmo neocortical é tão potente e flexível que com um pequeno reajuste, exclusivo dos humanos, é capaz de criar comportamentos novos e sofisticados. A outra resposta é que o comportamento e a predição são duas caras da mesma moeda. Ainda que o córtex possa imaginar o futuro, ele só

consegue realizar previsões sensoriais precisas se saber que comportamentos estão sendo postos em prática.

No simples exemplo do rato que busca o queijo, este recorda o labirinto e utiliza a memória para prever que verá o queijo ao virar a esquina. Mas o rato poderia girar à esquerda ou à direita; porém só se recordar ao mesmo tempo o queijo e o comportamento correto, "girar à direita na bifurcação", que ele será capaz de conseguir que a previsão sobre o queijo se torne realidade. Ainda que seja um exemplo trivial, ele serve para pôr de manifesto a íntima relação existente entre a previsão sensorial e o comportamento. Todo comportamento muda o que vemos, escutamos e sentimos. A maior parte do que sentimos em qualquer momento depende em grande parte das nossas próprias ações. Mova o braço diante do seu rosto. Para prever que verá seu braço, seu córtex tem que saber que você tem ordenado mover o braço. Se o córtex visse você mover-se seu braço sem a correspondente ordem motora, você se surpreenderia. O modo mais simples de interpretar isto seria supor que seu cérebro move primeiro o braço e depois prevê o que verá. Mas não é assim: o córtex prevê que verá o braço, e esta previsão é o que faz com que a ordem motora torne realidade a previsão. Você primeiro pensa, o que provoca que atue para conseguir que seus pensamentos se tornem realidade.

Vejamos agora as mudanças que conduziram os seres humanos a terem um repertório de comportamentos muito ampliado. Existem diferenças físicas entre o córtex de um macaco e a de um humano que possam explicar por que só o último desfruta da linguagem e outros comportamentos complexos? O cérebro humano é umas três vezes maior que o do chimpanzé. Mas não é só questão de tamanho. Uma chave para entender o salto no comportamento humano se encontra nas conexões entre regiões do córtex e partes do cérebro velho. Para expressar isto com maior simplicidade, nosso cérebro está ligado de forma diferente.

Vamos o observar com mais detalhamento. Todos nós conhecemos os hemisférios esquerdo e direito do cérebro. Mas existe outra divisão menos conhecida, que é a qual devemos analisar para buscar as diferenças humanas. Todos os cérebros, sobretudo os grandes, dividem o córtex em uma parte dianteira e outra traseira. Os cientistas empregam as palavras *anterior* para a metade frontal e *posterior* para a traseira. Um grande sulco, chamado cisura de Rolando, separa ambas metades. A parte posterior do córtex contém as seções nas quais chegam as entradas dos olhos, dos ouvidos e do tato. É onde ocorre em boa medida a percepção sensorial. A parte frontal contém regiões do córtex que participam no planejamento e no pensamento superiores. Também compreende o córtex motor, a seção do cérebro mais responsável pelo movimento dos músculos e, portanto, de criar o comportamento.

À medida que o neocórtex dos primatas foi-se alargando com o passar do tempo, a metade anterior atingiu um tamanho desproporcionado, sobretudo nos humanos.

Comparados com o restante dos primatas e dos primeiros homínídeos, temos frentes enormes projetadas para conter nosso grandíssimo córtex anterior. Mas este engrandecimento não basta para explicar a melhora da nossa capacidade motora comparada com a de outras criaturas. Nossa habilidade para efetuar movimentos excepcionalmente complexos provém do fato de que nosso córtex motor efetua muito mais conexões com os músculos dos nossos corpos. Em outros mamíferos o córtex anterior desempenha um papel menos direto no comportamento motor. A maioria dos animais dependem em boa medida das partes mais antigas do cérebro para gerar seu comportamento. Em contraste, o córtex humano usurpou a maior parte do controle motor do que o resto do cérebro. Se o córtex motor de um rato for lesado, é provável que ele não apresente carências apreciáveis. Se o córtex motor de um humano for lesado, ele ou ela ficará paralítico.

As pessoas costumam perguntar-me pelos golfinhos. Não têm eles cérebros enormes? Sim; o golfinho possui um neocórtex grande. Sua estrutura é mais simples (três camadas, em comparação às nossas seis) que a humana, mas em todos os parâmetros restantes é grande. É provável que o golfinho seja capaz de recordar e entender muitas coisas. Ele pode reconhecer a outros golfinhos particulares. É possível que conheça cada recanto do oceano onde tenha estado. Mas ainda que eles mostrem um comportamento um pouco sofisticado, os golfinhos não se aproximam ao nosso, pelo qual cabe conjecturar que seu córtex tem uma influência menos dominante sobre ele. A questão é que o córtex evoluiu primordialmente para proporcionar uma memória do mundo. Um animal com um córtex grande poderia perceber o mundo como você e eu. Mas os seres humanos são únicos quanto ao papel dominante e avançado que desempenha este em nosso comportamento. Por isso temos uma linguagem complexa e ferramentas refinadas, enquanto outros animais não; por isso podemos escrever novelas, navegar pela Internet, enviar sondas a Marte e construir navios de cruzeiro.

Agora podemos ver o quadro como um todo. A Natureza criou primeiro animais como os répteis com sentidos sofisticados e comportamentos complexos, mas relativamente rígidos. Depois descobriu que acrescentando um sistema de memória e lhe introduzindo o fluxo sensorial o animal podia recordar experiências passadas. Quando o animal se encontrava em uma situação igual ou parecida, a memória era recordada e se estabelecia uma predição do que era provável que acontecesse a seguir. Deste modo, a inteligência e o entendimento começaram como um sistema de memória que introduzia predições no fluxo sensorial. Estas predições são a essência do entendimento. Conhecer algo significa que você pode realizar predições a respeito.

O córtex evoluiu em duas direções. Primeiro fez-se maior e mais complexo nos tipos de memória que ele podia armazenar; ele era capaz de recordar mais coisas e realizar predições baseadas em relações mais complexas. Em segundo lugar, ele

começou a interagir com o sistema motor do cérebro velho. Para predizer o que se ia escutar, ver e sentir a seguir ele precisava saber que ações estavam sendo levadas a cabo. No caso dos humanos, o córtex tem assumido a maior parte do nosso comportamento motor. Em vez de limitar-se a realizar predições baseadas no comportamento do cérebro velho, o neocórtex humano dirige o comportamento para satisfazer suas predições.

O córtex humano é particularmente grande e, portanto, tem uma vasta capacidade de memória. Ele está predizendo de modo constante o que você verá, ouvirá e sentirá, ainda que na maioria das vezes não temos consciência disso. Estas predições são nossos pensamentos e, quando se combinam com as entradas sensoriais, nossas percepções. A esta visão do cérebro chamo-a de modelo de memória-predição da inteligência.

Se a sala chinesa de Searle contivesse um sistema de memória similar que pudesse realizar predições sobre que caracteres chineses apareceriam a seguir e o que aconteceria depois na história, estaríamos em situação de garantir que a sala entendia chinês e compreendia a história. Agora podemos ver que Alan Turing estava equivocado. A predição, e não o comportamento, é a prova da inteligência.

Já estamos preparados para aprofundar nos detalhes desta nova ideia sobre o modelo de memória-predição do cérebro. Para realizar predições sobre acontecimentos futuros, nosso neocórtex tem que armazenar sequências de padrões. Para recordar as memórias apropriadas ele tem que recuperar padrões por sua semelhança a padrões passados (lembrança auto-associativa). E, por último, as memórias têm de guardar em uma forma invariável a fim de que o conhecimento de acontecimentos passados seja aplicável a novas situações que são similares mas não idênticas ao passado. Como realiza estas tarefas o córtex, além da exploração mais completa da sua hierarquia, constituem o tema do próximo capítulo.

Como Funciona o Córtex

Tentar imaginar como funciona o cérebro é como resolver um quebra-cabeça gigantesco. Pode-se abordar de dois modos diferentes. Se utiliza-se a proposta de “acima-abaixo”, começa-se com a imagem que deve apresentar o quebra-cabeça resolvido, a empregando para decidir que peças devem ser ignoradas e que peças devem ser procuradas. Em outra proposta, de “abaixo-acima”, foca-se cada uma das peças por elas mesmas. Estudam-se suas características inusuais e busca-se como encaixar as mesmas com outras peças. Se não se conta com uma ilustração da solução do quebra-cabeça, o método de “abaixo-acima” é às vezes o único modo de proceder.

O quebra-cabeça de “compreender o cérebro” é particularmente intimidante. Ao carecer de um bom modelo para entender a inteligência, os cientistas viram-se obrigados a recorrer à proposta de “abaixo-acima”. Mas a tarefa é hercúlea, ou até impossível, com um quebra-cabeça tão complexo como o cérebro. Para se ter uma ideia da dificuldade, imagine um quebra-cabeça com vários milhares de peças. Muitas podem interpretar-se de múltiplos modos, como se cada uma tivesse uma imagem em ambas caras mas só uma fosse a correta. As peças mal apresentam formas, de modo que não se pode estar seguro se duas peças se encaixam ou não. Muitas delas não se utilizarão na solução definitiva, mas não se sabe quais ou quantas. A cada mês chegam novas peças por correio. Algumas substituem às antigas, como se o criador do quebra-cabeça dissesse: “Sei que está trabalhando há alguns dias com estas peças velhas, mas têm resultado estar erradas. Sinto muito. Utilize estas novas no seu lugar até futuro aviso”. Infelizmente, não se tem ideia de qual será o resultado final; e o que é pior, talvez se conte com algumas ideias, porém errôneas.

Esta analogia do quebra-cabeça é uma boa descrição da dificuldade que enfrentamos ao criar uma nova teoria do córtex e da inteligência. As peças do quebra-cabeça são os dados biológicos e comportamentais que os cientistas têm reunido durante mais de cem anos. A cada mês publicam-se novos artigos, que criam peças de quebra-cabeça adicionais. Às vezes os dados de um cientista contradizem os do outro. Como os ditos dados podem se interpretar de diferentes maneiras, existe desacordo sobre quase tudo. Sem uma estrutura de acima-abaixo não há consenso sobre o que buscamos, o que é mais importante ou como interpretar as montanhas de informação que se acumularam. Nosso entendimento do cérebro ficou preso na proposta de abaixo-acima. O que precisamos é de um modelo de acima-abaixo.

O modelo de memória-predição pode desempenhar esse papel. Pode-nos mostrar como começar a casar as peças do quebra-cabeça. Para realizar predições, o córtex precisa de um modo de memorizar e guardar o conhecimento sobre

sequências de acontecimentos. Para efetuar predições sobre acontecimentos novos, o córtex deve formar representações invariáveis. Nosso cérebro precisa criar e armazenar um modelo do mundo tal como é, independente de como o vemos em circunstâncias variáveis. Saber o que o córtex deve fazer nos guia para compreender sua arquitetura, sobretudo seu projeto hierárquico e seu formato de seis camadas.

À medida que vamos explorando este novo modelo, apresentado neste livro pela primeira vez, entrarei em um nível de detalhe que pode acabar sendo um desafio para alguns leitores. Muitos dos conceitos que estão a ponto de se encontrar são pouco conhecidos inclusive para os experts em neurociência. Mas com um pouco de esforço acho que qualquer pessoa pode entender os fundamentos desta nova estrutura. Os capítulos 7 e 8 são muito menos técnicos e exploram as repercussões mais amplas da teoria.

Nossa viagem para resolver o quebra-cabeça pode passar agora a buscar os detalhes biológicos que respaldam a hipótese da memória-predição; seria como conseguir separar uma grande porcentagem das peças do quebra-cabeça, sabendo que as relativamente poucas peças restantes vão revelar a solução definitiva. Uma vez que sabemos o que buscar, a tarefa se torna manejável.

Ao mesmo tempo, quero sublinhar que este novo modelo está incompleto. Há muitas coisas que ainda não compreendemos, mas há muitas outras que sim, nos baseando no raciocínio dedutivo, em experimentos realizados em muitos laboratórios diferentes, e na anatomia conhecida. Nos últimos cinco à dez anos, os pesquisadores de muitas subespecialidades da neurociência têm vindo explorando ideia similares à minha, embora empreguem uma terminologia diferente e não têm tratado, até onde sei, de colocar estas ideias em uma estrutura abrangedora. Eles falam de processamento de acima-abixo e abaixo-acima, de como os padrões se propagam pelas regiões sensoriais do cérebro e de como as representações invariáveis podem ser importantes. Por exemplo, os neurocientistas Caltech Gabriel Kreiman e Christof Koch, junto com o neurocirurgião Itzhak Fried, têm descoberto células que se estimulam sempre que uma pessoa vê uma foto de Bill Clinton. Uma das minhas metas é explicar como essas células do ex-presidente ganham vida. Certamente, todas as teorias precisam fazer predições que devem ser provadas em laboratório. Sugiro várias das ditas predições no apêndice. Agora que sabemos o que temos que buscar, este sistema não parecerá mais tão complexo.

Nas próximas seções deste capítulo iremos aprofundar no funcionamento do modelo de memória-predição. Começaremos com a estrutura e função em grande escala do neocórtex e avançaremos no entendimento das peças menores e como se encaixam no quadro geral.

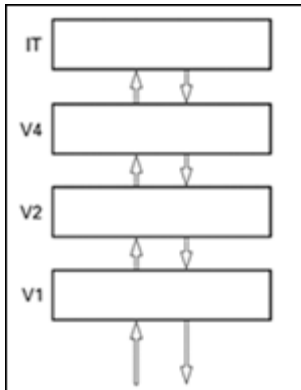


Figura 1. As quatro primeiras regiões visuais no reconhecimento de objetos.

Representações Invariáveis

Já apresentei a imagem do córtex como uma lâmina de células do tamanho de um guardanapo grande e da espessura de seis cartões de visita, na qual as conexões entre as diversas regiões outorgam ao todo uma estrutura hierárquica. Agora quero projetar outro quadro do córtex que destaque sua conectividade hierárquica. Imaginemos que cortamos o guardanapo em regiões funcionais diferentes — seções do córtex que se especializam em certas tarefas— e amontoamos as ditas regiões umas em cima das outras como se fossem panquecas. Se cortarmos esse monte e o virmos de lado, obteremos a figura 1. Preste atenção, o córtex não tem essa aparência na realidade, mas a imagem nos ajudará a visualizar como flui a informação. Tenho mostrado quatro regiões corticais nas quais a entrada sensorial entra por abaixo, na região inferior, e flui para acima de uma região a outra. Temos de ter em conta que a informação flui em ambos os sentidos.

A figura 1 representa as quatro primeiras regiões visuais que participam no reconhecimento de objetos: como se consegue ver e reconhecer um gato, uma catedral, a sua mãe, a Grande Muralha da China, o seu nome. Os biólogos chamam-nas de V1, V2, V4 e IT. A entrada visual representada pela seta inferior da figura origina-se nas retinas de ambos olhos e é transmitida a V1. Esta entrada pode conceber-se como os padrões em mudança constante transportados por aproximadamente um milhão de axônios que se unem para formar o nervo óptico.

Já falamos dos padrões espaciais e temporais, mas vale a pena refrescar a memória porque estarei me referindo a eles com frequência. Recordemos que o córtex é uma grande lâmina de tecido que contém áreas funcionais especializadas em determinadas tarefas. Estas regiões estão ligadas mediante grandes feixes de axônios ou fibras que transferem informação de uma região a outra tudo ao mesmo tempo. Em qualquer momento, um conjunto de fibras estimula impulsos elétricos, chamados potenciais de ação, enquanto outros permanecem quietos. A atividade coletiva em um feixe de fibras é o que se entende por padrão. O padrão

que chega a V1 pode ser espacial, como quando nossos olhos se fixam por um instante em um objeto, e temporal, como quando nossos olhos se movem pelo objeto.

Como já mencionei, umas três vezes por segundo nossos olhos efetuam um movimento rápido, chamado de sacada ocular, e uma parada, chamada de fixação. Se um cientista lhe equipasse com um aparelho que seguisse os movimentos oculares, você ficaria surpreso ao descobrir o quão entrecortadas são suas sacadas oculares, devido a você experimentar sua visão como se fosse contínua e estável. A figura 2a mostra como se moviam os olhos de uma pessoa enquanto olhava um rosto. Note que as fixações não são aleatórias. Agora imaginemos que podemos ver o padrão de atividade que chega a V1 desde os olhos dessa pessoa. Ele muda por completo a cada sacada ocular. Várias vezes por segundo o córtex visual vê um padrão completamente novo.

Caberia pensar: "Bem, mas continua sendo o mesmo rosto, só que mudado". É verdade até certo ponto, mas não tanto como se pensa. Os receptores de luz da nossa retina estão distribuídos de forma irregular. Concentram-se densamente no centro, chamado de fóvea, e vão se distribuindo pouco a pouco na periferia. O resultado é que a imagem retinal transmitida à área visual primária, V1, está muito distorcida. Quando nossos olhos se fixam no nariz de um rosto ao invés de um olho do mesmo rosto, a entrada visual é muito diferente, como se estivesse contemplando através de uma lente de olho de peixe distorcida que muda bruscamente de um lado a outro. Não obstante, quando vemos o rosto, não parece distorcido, nem também parece que esteja dando saltos. A maior parte do tempo nem sequer temos consciência de que o padrão da retina mudou, que dizer de forma tão dramática. Limitamo-nos a ver um "rosto". (A figura 2b mostra este efeito na visão de uma paisagem de praia.) É uma repetição do mistério da representação invariável da qual falamos no capítulo 4 sobre a memória. O que "percebemos" não é o que V1 vê. Como sabe nosso cérebro que está olhando o mesmo rosto e por que não nos damos conta de que as entradas estão mudando e estão distorcidas?

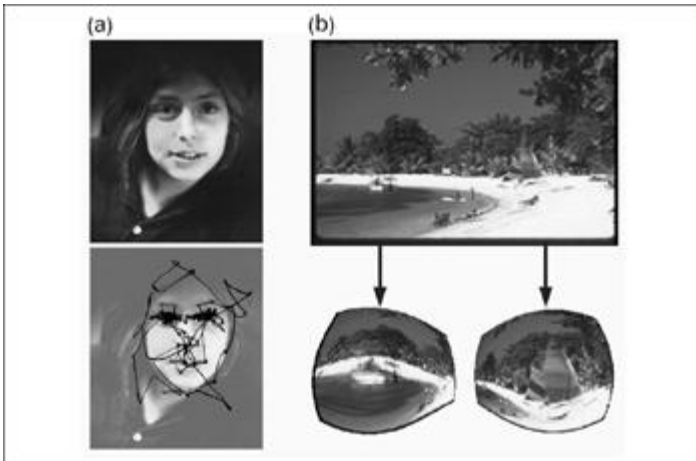


Figura 2a. Como o olho efetua sacadas oculares por um rosto humano. Figura 2b. Distorção causada pela distribuição irregular de receptores na retina.

Se colocamos uma sonda em V1 e observamos como respondem as células individuais, descobrimos que uma determinada célula só se estimula em resposta à entrada visual de uma diminuta parte da nossa retina. Este experimento realizou-se muitas vezes e é um dos suportes principais das pesquisas sobre a visão. Cada neurônio de V1 tem um assim chamado campo receptivo que é muito específico para uma parte diminuta do nosso campo de visão total, isto é, o mundo completo que há diante dos nossos olhos. As células de V1 parecem carecer de todo conhecimento sobre os rostos, carros, livros ou outros objetos significativos que vemos continuamente; não “conhecem” mais do que uma diminuta porção, do tamanho da ponta de um alfinete, do mundo visual.

Cada célula de V1 também está adaptada para tipos específicos de padrões de entrada. Por exemplo, uma célula particular poderia estimular-se muito quando visse uma linha ou uma borda inclinada a trinta graus dentro do seu campo receptivo. Essa borda tem pouco significado em si mesmo; poderia fazer parte de qualquer objeto, uma taboa de assoalho, o tronco de uma palmeira distante, a lateral de uma letra *M*, ou quaisquer outras possibilidades aparentemente infinitas. Com cada nova fixação, o campo receptivo da célula vai deter-se em uma porção nova e completamente diferente do espaço visual. Em algumas fixações a célula se estimulará; em outras se estimulará pouco ou não se estimulará. Portanto, cada vez que realizamos uma sacada ocular, é possível que muitas células de V1 mudem de atividade.

No entanto, algo mágico acontece se for colocada uma sonda na região superior que se mostra na figura 1, a região IT. Ali encontramos algumas células que se ativam e permanecem estimuladas quando aparecem objetos inteiros no campo visual. Por exemplo, poderíamos encontrar uma célula que se ativa com força sempre que é visível um rosto. Esta célula permanece ativa durante o tempo em

que nossos olhos olham um rosto em qualquer lugar do nosso campo de visão. Não se ativa e desativa com cada sacada ocular, como as células de V1. O campo receptivo desta célula de IT cobre o total do espaço visual e entra em atividade quando vê rostos.

Reconstruamos o mistério. No curso de abranger quatro estágios corticais da retina a IT, as células deixam de se dedicar ao reconhecimento de características diminutas que mudam rapidamente e apresentam uma especificidade espacial e passam a se especializar no reconhecimento de objetos, se estimulando de forma contínua e carecendo de especificidade espacial. A célula IT indica-nos que estamos vendo um rosto em qualquer lugar do nosso campo de visão. Esta célula, comumente chamada de célula de rosto, se estimulará não importando se o dito rosto estiver inclinado, rotacionado ou parcialmente tampado. Faz parte de uma representação invariável de rosto”.

Estas palavras escritas fazem com que isto pareça muito simples. Quatro rápidas etapas e voilá, reconhecemos um rosto. Nenhum programa de computador ou fórmula matemática tem resolvido este problema com nada que se aproxime à solidez e generalidade de um cérebro humano. Não obstante, sabemos que este o resolve em alguns passos, de modo que a resposta não pode ser tão difícil. Uma das metas primordiais deste capítulo é explicar como surge uma célula de rosto, seja o de Bill Clinton ou outro. Chegaremos a isso, mas antes devemos atender a muitos outros aspectos.

Demos outra olhada à figura 1. Podemos ver que a informação também flui das regiões inferiores às superiores por uma rede de conexões de realimentação. Trata-se de feixes de axônios que vão das regiões superiores como IT a regiões inferiores como V4, V2 e V1. Além disso, no córtex visual existem as mesmas conexões de realimentação —se não mais— que de alimentação.

Durante muitos anos, a maioria dos cientistas ignoraram estas conexões de realimentação. Se o entendimento deles do cérebro centrava-se em como o córtex aceitava as entradas, as processava e depois atuava em consequência, não precisavam de realimentação. Não era preciso mais do que conexões de alimentação que conduzissem das seções sensoriais às motoras do córtex. Mas quando começamos a nos dar conta de que a função central desta é realizar predições, temos que incorporar a realimentação ao modelo; o cérebro tem de enviar informação que reflua à região que recebe primeiro as entradas. A predição requer estabelecer uma comparação entre o que acontece e o que se espera que aconteça. O que acontece na realidade flui para acima e o que se espera que aconteça o faz para abaixo.

O mesmo processo de alimentação e realimentação ocorre em todas as áreas do córtex, abrangendo todos os nossos sentidos. A figura 3 mostra nosso monte de panquecas visual junto aos montes similares para a audição e o tato. Também

apresenta algumas regiões corticais superiores, às vezes chamadas de áreas de associação, que recebem e integram entradas de vários sentidos diferentes, como a audição o tato e a visão. Enquanto a figura 1 baseia-se na conectividade conhecida entre quatro regiões bem estudadas do córtex, a figura 3 não é mais do que um diagrama conceitual que não pretende captar as regiões corticais reais. Em um cérebro humano há dezenas de regiões corticais interconectadas de todo tipo de modos. De fato, a maior parte do córtex humano está composta por áreas de associação. A caracterização caricaturada que se mostra nesta e nas próximas figuras pretende lhe ajudar a compreender o que acontece sem confundir em nenhum aspecto importante.

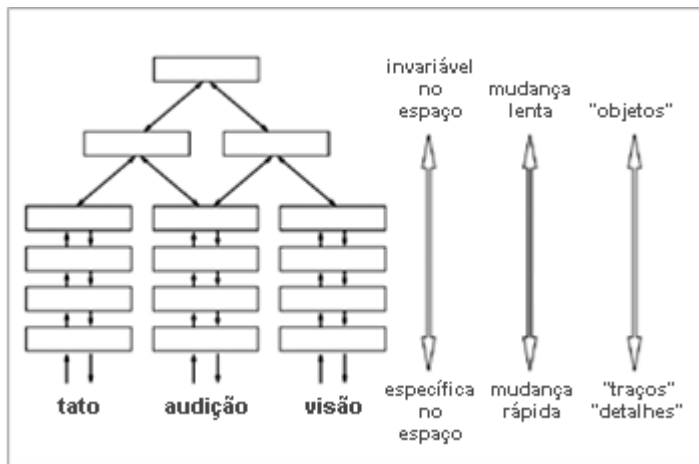


Figura 3. Formação de representações invariáveis na audição, na visão e no tato.

A transformação —de mudança rápida a lenta e de específica no espaço a invariável no espaço— está bem documentada para a visão. E ainda que exista um volume de provas menor para demonstrar isto, muitos neurocientistas acham que se descobrirá que o mesmo ocorre em todas as áreas sensoriais do córtex, não só na visão.

Tomemos a audição. Quando alguém te fala, as mudanças na pressão do som ocorrem com muita rapidez; os padrões que entram na área auditiva primária, chamada A1, também mudam igualmente rápido. No entanto, se conseguíssemos introduzir uma sonda mais acima no fluxo auditivo, poderíamos encontrar células invariáveis que respondem a palavras ou inclusive, em alguns casos, a expressões. Talvez nosso córtex auditivo possua um grupo de células que se estimulem quando escutamos “obrigado” e outro grupo quando escutamos a expressão “bom dia”. Estas células permaneceriam ativas enquanto durasse um enunciado, supondo que o reconheçamos.

Os padrões recebidos pela primeira área auditiva podem variar muito. Uma palavra pode ser dita com diferentes sotaques, em diferentes tons ou em velocidades

diversas. Porém mais acima, no córtex, essas características de baixo nível não importam; uma palavra é uma palavra independentemente dos detalhes acústicos. O mesmo cabe aplicar à música. Pode-se escutar *Três ratos cegos* tocada no piano, no clarinete ou cantada por uma criança, e a região A1 recebe um padrão completamente diferente em cada caso. Mas uma sonda colocada em uma região auditiva superior deveria encontrar células que se estimulassem de forma constante cada vez que se tocasse *Três ratos cegos* não importando o instrumento, compasso ou outros detalhes. Este experimento particular não se realizou, certamente, já que é muito agressivo para os seres humanos, mas se for aceito que deve existir um algoritmo cortical comum, pode-se presumir que as ditas células existem. Vemos o mesmo tipo de realimentação, predição e lembrança invariável no córtex auditivo que observamos no sistema visual.

Por último, o tato tem que se comportar da mesma maneira. Novamente, não se realizaram os experimentos definitivos, ainda que se esteja pesquisando com macacos em máquinas de imagens cerebrais de alta resolução. Enquanto estou sentado escrevendo, sustento uma caneta na mão. Toco a tampa e meus dedos acariciam seu gancho de metal. Os padrões que entram em meu córtex somatosensorial desde os sensores de tato da minha pele mudam com rapidez à medida que meus dedos se movem, porém percebo uma caneta constante. Em um dado momento poderia flexionar o gancho de metal com os dedos e em seguida fazer isto com dedos diferentes ou inclusive com os lábios. São entradas muito diferentes que chegam a localizações diferentes do córtex somatosensorial. No entanto, nossa sonda voltaria a encontrar células em regiões separadas a vários passos da entrada primária que respondem de forma invariável a "caneta". Permaneceriam ativas enquanto eu acariciasse a caneta e não se importariam com que dedos ou partes do meu corpo eu empregasse para tocá-la.

Pensemos nisto. Com a audição e o tato não se pode reconhecer um objeto com uma entrada momentânea. O padrão proveniente dos ouvidos ou os sensores de tato da pele não contém informação suficiente em nenhum ponto do tempo para indicar o que se está escutando ou sentindo. Quando percebemos uma série de padrões auditivos como uma melodia, uma palavra falada ou uma batida de porta, e quando percebemos um objeto tátil como uma caneta, o único modo de fazer é empregar o fluxo de entrada ao longo do tempo. Não podemos reconhecer uma melodia escutando uma nota, nem o tato de uma caneta com um único toque. Portanto, a atividade neural correspondente à percepção mental dos objetos, como as palavras faladas, deve durar mais no tempo que os padrões de entrada individuais. Este não é mais do que outro modo de chegar à mesma conclusão: quanto mais se sobe no córtex, menos mudanças devem ser vistas ao longo do tempo.

A visão também é um fluxo de entradas baseado no tempo e funciona do mesmo modo geral que a audição e o tato, mas como temos a faculdade de reconhecer

objetos particulares com uma única fixação, o quadro fica confuso. Efetivamente, esta capacidade de reconhecer padrões espaciais durante uma breve fixação tem desorientado durante muitos anos os pesquisadores que trabalham na visão das máquinas e dos animais. Em geral, eles têm passado por alto a natureza crucial do tempo. Ainda que em condições de laboratório pode-se fazer com que os seres humanos reconheçam objetos sem mover os olhos, não é a norma. A visão normal, como a sua ao ler este livro, requer um movimento ocular constante.

Integração dos Sentidos

*E o que acontece com as áreas de associação? Até agora temos visto subir e baixar os fluxos de informação em uma área sensorial particular do córtex. O fluxo descendente completa a entrada em curso e realiza predições sobre o que se experimentará a seguir. Ocorre o mesmo processo entre os sentidos —isto é, entre a visão, o som, o tato e muitos outros—. Por exemplo, algo que escuto pode levar a uma predição do que devo ver ou sentir. Agora estou escrevendo no meu quarto. Nossa gata *Keo* tem um colar que chocalha ela quando caminha. Escuto seu chocalhar aproximar desde o corredor. Por esta entrada auditiva reconheço a minha gata; viro a cabeça para o corredor e aparece *Keo*. Espero vê-la baseando no seu som. Se ela não entra, ou aparece outro animal, eu ficaria surpreso. Neste exemplo, uma entrada auditiva criou primeiro um reconhecimento auditivo de *Keo*. A informação fluiu para acima na hierarquia auditiva até uma área de associação que liga a visão com a audição. Então a representação voltou a fluir para abaixo nas hierarquias auditiva e visual, conduzindo a predições auditivas e visuais. A figura 4 ilustra isto.*

Este tipo de predição multisensorial ocorre continuamente. Torço para fora o gancho da minha caneta, sinto como ele se solta dos meus dedos e espero escutar um estalo quando ele golpeia o canhão da tampa. Se não escutasse o estalo após soltar o gancho, eu ficaria surpreso. Meu cérebro prediz com precisão quando escutarei o som e como será. Para que ocorra esta predição, a informação tem fluído para acima pelo córtex somatosensorial e novamente para abaixo pelo córtex somatosensorial e auditivo, levando à predição de que escutaria e sentiria um estalo.

Outro exemplo: vou de bicicleta para o trabalho em vários dias da semana. Nessas manhãs entro na garagem, apanho a bicicleta, dou-lhe a volta e saio a empurrando até a porta. No processo recebo muitas entradas visuais, táteis e auditivas. A bicicleta golpeia a batente da porta, a corrente matraqueia, um pedal encosta na minha perna e a roda gira quando fricciona o solo. No processo de transportar a bicicleta para fora da garagem, meu cérebro topa com uma enxurrada de sensações de visão, som e tato. Cada fluxo de entrada realiza predições para as demais de um modo tão coordenado que surpreende. As coisas que vejo conduzem a predições precisas sobre as coisas que sinto e escuto, e vice-versa. Ver como a bicicleta golpeia a batente da porta faz com que eu espere

escutar um som particular e sentir ela saltar para cima. Sentir que o pedal encosta na minha perna me incita a olhar para abaixo e predizer que verei o pedal justo onde o sinto. As predições são tão precisas que me daria conta se alguma destas entradas estivesse levemente descoordenada ou não fosse usual. A informação flui de modo simultâneo para acima e para abaixo das hierarquias sensoriais para criar uma experiência sensorial unificada que envolve predição em todos os sentidos.

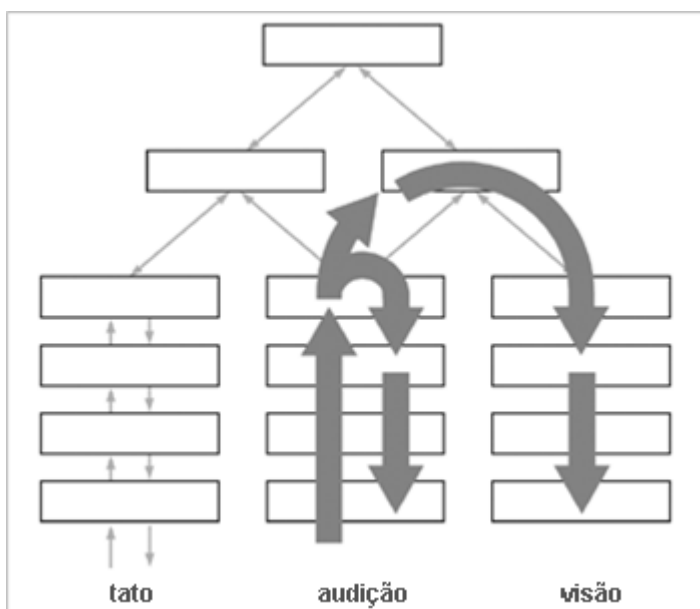


Figura 4. A informação flui acima e abaixo nas hierarquias sensoriais para formar predições e criar uma experiência sensorial unificada.

Tente este experimento. Deixe de ler e faça algo, uma atividade que envolva mover o corpo e manipular um objeto. Por exemplo, vá à pia e abra a torneira. Agora, enquanto o faz, tente perceber todo som, sensação de tato e entradas visuais mutáveis. Terá que se concentrar. Cada ação está intimamente unida com visões, sons e sensações de tato. Levante ou gire a alavanca da torneira: seu cérebro espera sentir pressão sobre a pele e resistência nos músculos. Você espera ver e sentir se mover a alavanca, e ver e escutar a água. Quando a água cai na pia, você espera escutar um som diferente e ver e sentir o respingo.

Todas as pisadas produzem um som que você sempre antecipa, tendo consciência disso ou não. Inclusive o simples ato de segurar este livro leva a muitas predições sensoriais. Imagine que sentisse e escutasse o livro se fechar, mas visualmente ele permanecesse aberto. Você ficaria surpreso e confuso. Como vimos no caso da porta modificada no experimento apresentado no capítulo 5, fazemos predições constantes do mundo coordenadas com todos os sentidos. Quando me concentro em todas as pequenas sensações, me assombra o quão plenamente integradas estão nossas predições perceptivas. Ainda que estas percepções possam parecer simples ou triviais, não há por que esquecer seu caráter omnipresente e que só

podem ocorrer mediante a coordenação dos padrões que fluem em ambos sentidos da hierarquia cortical.

Uma vez que se compreende a grande interconexão dos sentidos, se chega à conclusão de que o neocórtex por inteiro, e todas as áreas sensoriais e de associação, atuam como um só sistema. Sim, temos um córtex visual, mas não é mais do que um componente de um sistema sensorial único e abrangente: visões, sons, tatos e todo o demais combinado fluindo acima e abaixo de uma única hierarquia com múltiplos ramos.

Mais uma questão: todas as predições se aprendem por experiência. Esperamos que os ganchos das canetas façam determinados sons no presente e no futuro porque os fizeram no passado. As bicicletas que se golpeiam nas garagens são vistas, sentidas e ouvidas de modos predizíveis. Não nascemos com nenhum destes conhecimentos; os aprendemos graças à grandíssima capacidade do nosso córtex para recordar padrões. Se há padrões constantes entre as entradas que fluem a nossos cérebros, nosso córtex os usará para predizer acontecimentos futuros.

Ainda que as figuras 3 e 4 não representem o córtex motor, você pode o imaginar como outro monte hierárquico de panquecas, semelhante ao monte sensorial, ligado aos sistemas sensoriais pelas áreas de associação (talvez com conexões mais íntimas com o córtex somatosensorial para efetuar movimentos corporais). Deste modo, o córtex motor comporta-se quase da mesma forma que uma região sensorial. Uma entrada em uma área sensorial pode fluir para acima até uma área de associação, o que pode fazer com que um padrão flua para abaixo ao córtex motor e se obtenha como resultado um comportamento. Do mesmo modo que uma entrada visual pode fazer com que os padrões fluam para abaixo às seções auditiva e tátil do córtex, também pode fazer com que um padrão flua para abaixo à seção motora do córtex. No primeiro caso, interpretamos esses padrões que fluem para abaixo como ordens motoras. Como assinalou Mountcastle, o córtex motor se parece ao córtex sensorial. Portanto, sua forma de processar predições sensoriais que fluem para abaixo é similar ao seu modo de processar ordens motoras que fluem no mesmo sentido.

Veremos cedo que não existem áreas puramente sensoriais ou motoras no córtex. Os padrões sensoriais fluem a todas partes, e depois voltam a descer por uma área da hierarquia, levando a predições ou comportamentos motores. Ainda que o córtex motor possua alguns atributos especiais, é razoável concebê-lo como parte de um grande sistema de memória-predição hierárquico. É quase como outro sentido. Ver, ouvir, tocar e agir estão profundamente entrelaçados.

Uma Nova Visão de V1

O próximo passo para desvendar a arquitetura do córtex requer que olhemos todas as regiões corticais de uma nova maneira. Sabemos que as regiões superiores da hierarquia cortical formam representações invariáveis, mas por que esta importante função só ocorre no topo? Tendo em mente a noção de simetria de Mountcastle, comecei a explorar os modos diferentes nas quais as regiões corticais poderiam estar ligadas.

A figura 1 descreve as quatro regiões clássicas da rota visual, V1, V2, V4 e IT, com V1 na base do monte, V2 e V4 acima, e IT no topo. Costuma-se considerar e mostrar cada região como se fosse única e contínua. Portanto, supõe-se que todas as células de V1 fazem coisas similares, ainda que com diferentes partes do campo visual. Todas as células de V2 efetuam o mesmo tipo de tarefas. Todas as células de V4 possuem uma especialização semelhante.

Nesta proposta tradicional, quando a imagem de um rosto entra na região de V1, as células que estão nela criam um esboço tosco dele se servindo de simples segmentos de linha e outras características elementares. O esboço é enviado a V2, onde a imagem recebe uma análise um pouco mais sofisticada das características faciais, que a seguir é enviada a V4, e assim sucessivamente. A invariância e o reconhecimento do rosto só são obtidas quando a entrada chega ao topo, IT.

Infelizmente, esta proposta uniforme das primeiras regiões corticais como V1, V2 e V4 apresenta alguns problemas. Mais uma vez, por que só devem aparecer representações invariáveis em IT? Se todas as regiões do córtex realizam a mesma função, por que IT tem de ser especial?

Em segundo lugar, o rosto pode aparecer no lado esquerdo ou direito do nosso V1 e o reconheceríamos. Mas os experimentos mostram com clareza que as partes não adjacentes de V1 não estão diretamente ligadas; o lado esquerdo de V1 não pode saber de forma direta que está vendo o lado direito. Voltemos um pouco para pensar sobre isso. Sem dúvida, as diferentes partes de V1 estão fazendo algo similar, já que todas podem participar no reconhecimento de um rosto, mas ao mesmo tempo são fisicamente independentes. As sub-regiões ou grupos de V1 estão desligadas desde o ponto de vista físico, mas fazem o mesmo.

Por último, os experimentos mostram que todas as regiões superiores do córtex recebem entradas convergentes de duas ou mais regiões sensoriais abaixo (figura 3). Nos cérebros reais uma dúzia de regiões podem convergir em uma área de associação. Mas nas representações tradicionais as regiões sensoriais inferiores como V1, V2 e V4 parecem ter um tipo diferente de conectividade. Parece como se cada uma não tivesse mais do que uma única fonte de entradas —só uma seta flui para cima desde o fundo— e não há uma convergência clara de entradas desde as diferentes regiões. V2 obtém suas entradas de V1, e isso é tudo. Por que

algumas regiões corticais recebem entradas convergentes e outras não? Isto também não concorda com a ideia de Mountcastle do algoritmo cortical comum.

Por esta e outras razões, tenho chegado a pensar que V1, V2 e V4 não devem se considerar regiões corticais únicas, mas cada qual uma reunião de muitas sub-regiões menores. Voltemos à analogia do guardanapo grande, que constitui uma versão achatada do córtex como um todo. Digamos que fôssemos utilizar uma caneta para marcar todas as regiões funcionais do córtex em nosso guardanapo cortical. A região maior significativamente é V1, a área visual primária. A seguir seria V2. São enormes comparadas com a maioria das regiões. O que sugiro é que V1 deve ser considerada na realidade muitas regiões muito pequenas. Em vez de uma grande área do guardanapo, projetaríamos muitas áreas pequenas que juntas ocupariam a área que costuma se atribuir a V1. Em outras palavras, V1 está composta por numerosas pequenas áreas corticais separadas que só estão ligadas com suas vizinhas de forma indireta através de regiões que se encontram acima na hierarquia. V1 teria um maior número de pequenas sub-regiões que qualquer área visual. V2 também estaria composta por sub-regiões, ainda que em menor número e um pouco maiores. O mesmo se aplica a V4. Mas quando chegamos à região superior, IT, nós encontramos realmente uma região ímpar, motivo pelo qual suas células gozam de uma vista aérea do mundo visual como um todo.

Existe uma agradável simetria. Demos uma olhada à figura 5, que mostra a mesma hierarquia que a figura 3, porém refletindo as hierarquias sensoriais como as acabo de descrever. Note que agora o córtex parece similar em todas as partes. Escolha uma região e encontrará muitas regiões inferiores que proporcionam entradas sensoriais convergentes. A região receptora reenvia projeções a suas regiões de entrada, indicando-lhes que padrões devem esperar ver a seguir. As áreas de associação superiores unem informação proveniente de múltiplos sentidos como a visão e o tato. Uma região inferior como V2 une a informação de sub-regiões separadas dentro de V1. Uma região não sabe —nem pode saber— o que significam quaisquer dessas entradas. Uma sub-região de V2 não precisa saber que está manejando entradas visuais procedentes de múltiplas partes de V1. Uma área de associação não precisa saber que está manejando entradas da visão e da audição. Ao invés disso, a tarefa de qualquer região cortical é descobrir como se relacionam suas entradas, memorizar a sequência de correlações entre elas e utilizar esta memória para prever como se comportarão as entradas no futuro. O córtex é o córtex. O mesmo processo ocorre em todas as partes: um algoritmo cortical comum.

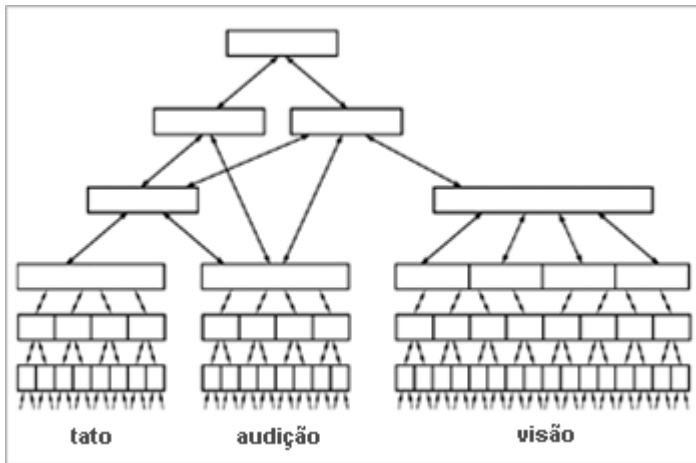


Figura 5. Proposta alternativa da hierarquia cortical

Esta nova representação hierárquica ajuda-nos a compreender o processo de criação de representações invariáveis. Observemos com maior detalhamento como funciona na visão. No primeiro nível de processamento, o lado esquerdo do espaço visual é diferente do direito, do mesmo modo que a audição é diferente da visão. V1 esquerdo e V1 direito formam o mesmo tipo de representações somente por que são expostos a padrões similares na vida. Assim como a audição e a visão, cabe os considerar fluxos sensoriais separados que se unem em planos superiores.

Da mesma forma, as regiões menores dentro de V2 e V4 são áreas de associação da visão. (As sub-regiões podem se sobrepor, o que não mudaria em essência sua forma de funcionar.) Interpretar o córtex visual desta maneira não contradiz nem muda nada do que conhecemos sobre sua anatomia. A informação flui para acima e para abaixo em todos os ramos da árvore de memória hierárquica. Um padrão do campo visual esquerdo pode levar a uma predição no campo visual direito, do mesmo modo que o chocalho da minha gata pode incitar a predição visual de que ela está entrando em meu quarto.

O resultado mais importante desta nova representação da hierarquia cortical é que agora podemos afirmar que todas e cada uma das regiões corticais formam representações invariáveis. Na proposta antiga não contávamos com representações invariáveis completas —como rostos— até que as entradas atingiam a camada suprema, IT, que vê o mundo visual completo. Agora podemos afirmar que as representações invariáveis são onipresentes; formam-se em todas as regiões corticais. A invariância não é algo que aparece de forma mágica quando se atingem as regiões superiores do córtex, como IT. Cada região forma representações invariáveis extraídas da área de entrada inferior desde a perspectiva hierárquica. Assim, as sub-regiões de V4, V2 e IT criam representações invariáveis baseadas no que flui a elas. É provável que só vejam uma parte diminuta do mundo e que o vocabulário de objetos sensoriais que manejem seja mais básico, mas realizam a mesma tarefa que IT. Além disso, as regiões de

associação acima de IT formam representações invariáveis de padrões procedentes de múltiplos sentidos. Deste modo, todas as regiões do córtex formam representações invariáveis do mundo que têm abaixo na hierarquia. É algo maravilhoso.

Nosso quebra-cabeça mudou. Já não temos que nos perguntar como se formam as representações invariáveis em quatro passos de abaixo-acima, mas como se formam em cada uma das regiões corticais, o que tem muito sentido se nós levamos a sério a existência de um algoritmo cortical comum. Se uma região armazena sequências de padrões, todas as regiões criam representações invariáveis. A reformulação da hierarquia cortical segundo as linhas mostradas na figura 5 possibilita esta interpretação.

Um Modelo do Mundo

Por que o neocórtex está construído como uma hierarquia?

Podemos pensar sobre o mundo, mover-nos por ele e efetuar predições sobre o futuro porque nosso córtex tem criado um modelo do mundo. Um dos conceitos mais importantes deste livro é que sua estrutura hierárquica guarda um modelo da estrutura hierárquica do mundo real. A estrutura aninhada do mundo real reflete-se na estrutura aninhada do nosso córtex.

O que entendo por estrutura aninhada ou hierárquica? Pensemos na música. As notas combinam-se para formar intervalos; os intervalos combinam-se para formar frases melódicas; as frases combinam-se para formar melodias ou canções; as canções combinam-se em álbuns. Pensemos na linguagem escrita. As letras combinam-se para formar sílabas; as sílabas combinam-se para formar palavras; as palavras combinam-se para formar frases e orações. Para abordar de um modo totalmente diferente, pensemos em nossos bairros. É provável que contenham estradas, colégios e casas. As casas têm cômodos. Cada cômodo tem paredes, teto, piso, porta e uma ou mais janelas. Cada um desses elementos está composto de objetos menores. As janelas são feitas de vidro, marcos, fechaduras e persianas. As fechaduras compõem-se de peças menores, como os parafusos.

Dediquemos um momento a observar nosso ambiente. Os padrões da retina que entram em nosso córtex visual primário combinam-se para formar segmentos de linha. Os segmentos de linha combinam-se para criar formas mais complexas. Estas formas mais complexas combinam-se para formar objetos como os narizes. Os narizes combinam-se com os olhos e com as bocas para formar rostos. E os rostos combinam-se com outras partes do corpo para formar a pessoa que está sentada na sala em frente a você.

Todos os objetos do nosso mundo estão compostos por subobjetos que aparecem sempre juntos; essa é a mera definição de objeto. Quando atribuímos um nome a

algo, o fazemos porque há um conjunto de características que estão constantemente juntas. Um rosto é um rosto por que sempre aparecem juntos dois olhos, um nariz e uma boca. Um olho é um olho porque sempre aparecem juntos uma pupila, uma íris, uma pálpebra e assim por diante. O mesmo cabe afirmar das cadeiras, dos carros, das árvores, dos parques e dos países. E, para finalizar, uma canção é uma canção porque sempre aparecem juntos em uma sequência uma série de intervalos.

Deste modo, o mundo é como uma canção. Cada objeto do mundo está composto por uma reunião de objetos menores, e a maioria dos objetos fazem parte de outros maiores. Isto é o que entendo por estrutura aninhada. Uma vez que temos consciência dela, podemos ver uma estrutura aninhada em todas partes. De modo análogo, nossas lembranças das coisas e a forma como nosso cérebro as representa se guardam na estrutura hierárquica do córtex. A memória da nossa casa não existe em uma região do córtex. Guarda-se em uma hierarquia de regiões corticais que reflete a estrutura hierárquica da casa. As relações de grande escala armazenam-se no topo da hierarquia, e as relações de pequena escala, no fundo.

O projeto do córtex e o método pelo qual aprende descobrem de forma natural as relações hierárquicas do mundo. Não nascemos com o conhecimento da linguagem, das casas ou da música. O córtex possui um algoritmo de aprendizagem inteligente que descobre e capta de forma natural qualquer estrutura hierárquica que exista. Quando falta a dita estrutura, caímos na confusão e inclusive no caos.

Em um determinado momento, só somos capazes de experimentar um subconjunto do mundo. Só podemos estar em um cômodo da nossa casa, olhando em uma direção. Devido à hierarquia do córtex, conseguimos saber que estamos em casa, na nossa sala, olhando uma janela, ainda que nesse momento dê a casualidade de que nossos olhos se tenham fixado no trinco da janela. As regiões superiores do córtex mantêm uma representação da nossa casa, enquanto as inferiores representam os cômodos, e outras mais inferiores olham a janela. De modo similar, a hierarquia permite-nos saber que estamos escutando uma canção e um álbum de música, ainda que em um determinado momento só escutemos uma nota, que por si mesma não nos diz quase nada. Permite-nos saber que estamos com nossa melhor amiga, ainda que nossos olhos se tenham fixado momentaneamente em sua mão. As regiões superiores do córtex estão atentas ao quadro geral, enquanto as áreas inferiores ocupam-se dos pequenos detalhes que mudam com rapidez.

Já que só podemos tocar, escutar e ver partes muito pequenas do mundo em um determinado momento do tempo, a informação que flui ao cérebro chega de forma natural como uma sequência de padrões. O córtex quer aprender essas sequências que aparecem uma e outra vez. Em alguns casos, como no das melodias, uma sequência de padrões chega em uma ordem rígida, ou seja, a ordem dos

intervalos. A maioria de nós conhecemos esse tipo de sequência. Mas vou utilizar a palavra sequência de um modo mais geral, mais próximo em significado ao termo matemático conjunto. Uma sequência é um conjunto de padrões que costumam se acompanhar mutuamente, mas nem sempre em uma ordem fixa. O importante é que padrões de uma sequência se seguem uns a outros no tempo, ainda que não seja em uma ordem fixa.

Alguns exemplos clarearão este ponto. Quando olho seu rosto, a sequência de padrões de entrada que vejo não é fixa, pois está determinada por minhas sacadas oculares. Em um momento eu poderia fixar na ordem "olho olho nariz boca" e logo depois fixar na ordem "boca olho nariz olho". Os componentes de um rosto são uma sequência. Estão relacionados estatisticamente e tendem a aparecer juntos no tempo, ainda que a ordem possa variar. Se percebemos "rosto" enquanto fixamo-nos em "nariz", os próximos padrões prováveis seriam "olho" ou "boca", mas não "caneta" ou "carro".

Cada região do córtex vê um fluxo dos ditos padrões. Se estão relacionados de um modo que permite à região aprender a prever que padrão aparecerá a seguir, a região cortical formará uma representação persistente ou uma memória para a sequência. Aprender sequências é o requisito básico para formar representações invariáveis dos objetos do mundo real.

A possibilidade de predição é a definição da realidade. Se uma região do córtex descobre que pode se mover com fidelidade e previsão entre estes padrões de entrada utilizando uma série de movimentos físicos (como as sacadas oculares dos olhos ou as caricias dos dedos) e os prever com precisão à medida que se desdobram no tempo (como os sons que envolvem uma canção ou uma palavra falada), o cérebro interpreta que eles apresentam uma relação causal. A probabilidade de que apareçam numerosos padrões de entrada na mesma relação uma e outra vez por pura coincidência é pequeníssima. Uma sequência predizível de padrões faz parte de um objeto maior que existe realmente. Portanto, a possibilidade de predição confiável é um modo seguro de saber que diferentes acontecimentos do mundo estão unidos fisicamente. Todos os rostos têm olhos, orelhas, boca e nariz. Se o cérebro vê um olho, efetua uma sacada ocular e vê outro olho, efetua mais uma sacada ocular e vê uma boca, ele pode se sentir seguro de que está contemplando um rosto.

Se as regiões corticais pudessem falar, talvez dissessem: "Experimento diferentes padrões. Às vezes não sou capaz de prever que padrão verei a seguir. Mas esses conjuntos de padrões estão relacionados entre si. Sempre aparecem juntos e posso saltar confiantemente entre eles. De modo que sempre que eu ver um desses eventos me referirei a eles com um nome comum". É este nome de grupo que vai ser passado às regiões superiores do córtex.

Portanto, cabe afirmar que o cérebro armazena sequências de sequências. Cada região do córtex aprende sequências, desenvolve o que chamo de “nomes” para as sequências que conhece e passa esses nomes às próximas regiões superiores da hierarquia cortical.

Sequências de sequências

À medida que a informação avança para cima desde as regiões sensoriais primárias até os níveis superiores, vemos cada vez menos mudanças ao longo do tempo. Nas áreas visuais primárias como V1, o conjunto de células ativas muda com rapidez à medida que entram na retina novos padrões várias vezes por segundo. Na área visual IT, os padrões de ativação das células são mais estáveis. O que acontece lá? Cada região do córtex possui um repertório de sequências que conhece, análogo a um repertório de canções. As regiões armazenam essas sequências semelhantes a canções sobre todas e cada uma das coisas: o som das ondas que quebram na praia, o rosto da sua mãe, o caminho da sua casa à loja da esquina, a ortografia da palavra “pipoca”, ou como embaralhar as cartas.

De forma similar aos nomes que temos para as canções, cada região cortical possui um nome para cada sequência que conhece. Este “nome” é um grupo de células cuja ativação coletiva representa o conjunto de objetos da sequência. (Não importa por agora como se seleciona esse grupo de células para representar a sequência; nos ocuparemos disso mais adiante.) Estas células permanecem ativas enquanto a sequência é representada, e é seu “nome” que é passado à próxima região da hierarquia. Enquanto os padrões de entrada fazem parte de uma sequência predizível, a região apresenta um “nome” constante às próximas regiões superiores.

É como se a região estivesse dizendo: “Aqui está o nome da sequência que estou escutando, vendo ou tocando. Você não precisa conhecer as notas, bordas ou texturas individuais. Te farei saber se algo novo imprevisível acontece”. Mais em concreto, cabe imaginar a região IT do topo da hierarquia visual se comunicando com uma área de associação abaixo: “Estou vendo um rosto. Sim, com cada sacada ocular os olhos fixam-se em partes diferentes do rosto; estou vendo partes diferentes do rosto em sucessão. Mas continua sendo o mesmo rosto. Te comunicarei quando eu ver algo mais”. Deste modo, uma sequência predizível de acontecimentos fica identificada com um “nome”, um padrão constante de ativação celular. Isto ocorre uma e outra vez enquanto ascendemos na pirâmide hierárquica. Uma região poderia reconhecer uma sequência de sons que compreende fonemas (os sons que constituem as palavras) e passar um padrão que representa o fonema à próxima região superior. Esta reconhece a sequência de fonemas para criar palavras. A próxima região superior reconhece sequências de palavras para criar orações, e assim sucessivamente. Não se deve esquecer que uma “sequência” nas regiões mais baixas do córtex pode ser muito simples, como uma borda visual que se move pelo espaço.

Ao desintegrar sequências predizíveis em objetos com “nome” em cada região da nossa hierarquia, conseguimos a cada vez maior estabilidade à medida que ascendemos, e assim se criam as representações invariáveis.

Acontece o efeito contrário quando um padrão desce na hierarquia: os padrões estáveis se “desdobram” em sequências. Suponhamos que memorizamos o Discurso de Gettysburg quando estávamos na sétima série e agora queremos recitá-lo. Em uma região de linguagem superior do nosso córtex existe um padrão guardado que representa o famoso discurso oficial de Lincoln. Primeiro este padrão se desdobra em uma memória da sequência das orações. Na próxima região inferior, cada frase se desdobra em uma memória da sequência das palavras. Neste ponto, o padrão despregado divide-se e viaja para abaixo tanto à seção auditiva como à seção motora do córtex. Seguindo a rota motora, cada palavra se desdobra em uma sequência memorizada de fonemas. E na região final inferior, cada fonema se desdobra em uma sequência de ordens musculares para criar os sons. Quanto mais abaixo olharmos na hierarquia, com maior rapidez mudarão os padrões. Um padrão único e constante do topo da hierarquia motora acaba levando a uma sequência complexa e prolongada de sons da fala.

A invariância também funciona a nosso favor quando esta informação desce na hierarquia. Se queremos digitar o Discurso de Gettysburg em vez de pronunciar-lo, começamos com o mesmo padrão no topo da hierarquia. O padrão se desdobra em orações na região abaixo. As orações se desdobram em palavras na região inferior. Até agora não há diferença entre pronunciar e digitar o discurso. Mas no próximo nível inferior nosso córtex motor toma uma rota diferente. As palavras se desdobram em letras, e estas, em ordens musculares a nossos dedos para que digitem: “Há oitenta e sete anos, nossos pais fundaram...”. As memórias das palavras são manejadas como representações invariáveis; não importa se as vamos pronunciar, as digitar ou as escrever a mão. Note que não temos que memorizar o discurso duas vezes, uma para o pronunciar e outra para o escrever. Uma única memória pode adotar várias formas de comportamento. Em qualquer região, um padrão invariável pode bifurcar-se e seguir uma rota descendente diferente.

Em uma mostra complementar de eficiência, as representações de objetos simples no fundo da hierarquia podem voltar a ser utilizadas uma ou outra vez para sequências diferentes de nível superior. Por exemplo, não temos que aprender um conjunto de palavras para o Discurso de Gettysburg e outro completamente diferente para “Eu tenho um sonho” de Martin Luther King, ainda que as duas orações contenham algumas palavras iguais. Uma hierarquia de sequências aninhada permite compartilhar e reutilizar objetos de nível inferior; palavras, fonemas e letras não são mais do que um exemplo. É uma forma muito eficiente de armazenar informação sobre o mundo e sua estrutura, muito diferente do modo como funcionam os computadores.

O mesmo desdobre de sequências ocorre nas regiões sensoriais e motoras. O processo permite-nos perceber e compreender objetos de perspectivas diferentes. Se você caminha para sua geladeira para apanhar um sorvete, seu córtex visual é ativado em múltiplos planos. Em um nível superior percebe uma “geladeira” constante. Em regiões inferiores esta expectativa visual rompe-se em uma série de entradas visuais mais localizadas. Ver a geladeira é composto de fixações no puxador da porta, no dispensador de gelo, nos ímãs sobre a porta, em um desenho infantil e assim por diante. Nos poucos milésimos de segundo que se passam enquanto efetuamos uma sacada ocular de uma característica a outra da geladeira, descem em sucessão as predições sobre o resultado de cada sacada ocular por nossa hierarquia visual. Enquanto as ditas predições confirmam-se uma sacada ocular após outra, nossas regiões visuais superiores continuam satisfeitas com que está vendo na realidade em sua geladeira. Notemos que neste caso, diferente da ordem fixa de palavras do Discurso de Gettysburg, a sequência que você vê quando olha o frigorífico não está fixa; o fluxo de entradas e os padrões da memória recuperada dependem das suas próprias ações. Portanto, em um caso como este, o padrão que se desdobra não é uma sequência rígida, mas o resultado final é o mesmo: padrões de nível superior que mudam com lentidão, se despregando em padrões de nível inferior que mudam com rapidez.

Nossa forma de memorizar sequências e representá-las com um nome enquanto a informação ascende e desce por nossa hierarquia cortical talvez lhe recorde a hierarquia de comando militar. O general supremo do exército diz: “Transfira as tropas para Flórida para passar o inverno”. Esta simples ordem de alto nível se desdobra em sequências cada vez mais detalhadas de ordens à medida que vai descendo na hierarquia. Os subalternos do general reconhecem que a ordem requer uma sequência de passos, como os preparativos para se marchar, o transporte para Flórida e os preparativos da chegada. Cada um destes passos se descompõe em outros mais específicos nos quais devem realizar os subordinados. Na base há milhares de soldados rasos efetuando centenas de milhares de ações que dão como resultado a transferência das tropas. Em cada nível são gerados relatórios do que tem acontecido. À medida que vão ascendendo na hierarquia, vão-se resumindo uma e outra vez, até que no topo o general recebe uma informação diária que diz: “A transferência para Flórida foi bem sucedida”. O general não obtém todos os detalhes.

Existe uma exceção a esta regra. Se algo vai mau e não o podem solucionar os subordinados inferiores da corrente de comando, o assunto vai ascendendo na hierarquia até que alguém saiba o que deve ser feito a seguir. O oficial que sabe como manejar a situação não a considera uma exceção. O que para os subordinados constituía um problema imprevisto não é mais do que a próxima tarefa esperada em uma lista. Então o oficial emite novas ordens a seus subordinados. O neocórtex comporta-se de modo similar. Como veremos daqui a pouco, quando ocorrem acontecimentos (em outras palavras, padrões)

imprevistos, a informação a respeito ascende na hierarquia cortical até que alguma região a possa manejar. Se as regiões inferiores não conseguem prever os padrões que estão vindo, consideram isto como um erro e o passam para acima da hierarquia. Isto se repete até que alguma região preveja o padrão.

* * *

Por seu projeto, cada região cortical tenta armazenar e recordar sequências. Mas esta continua sendo uma descrição muito simples do cérebro. Precisamos acrescentar mais algumas complexidades ao modelo.

As entradas de abaixo-acima a uma região do córtex são padrões de entrada transportados em bilhões de axônios. Estes axônios provem de diferentes regiões e contêm todo o tipo de padrões. O número de padrões possíveis que podem existir inclusive em um milhão de axônios é maior do que o número de moléculas do Universo. Uma região só verá uma diminuta fração dos ditos padrões em toda sua vida.

Portanto, propõe-se uma pergunta: quando uma única região armazena sequências, de que são? A resposta é que uma região classifica primeiro suas entradas como uma de um número limitado de possibilidades e depois busca sequências. Imaginemos que somos uma determinada região cortical. Nossa tarefa é classificar pedaços de papel coloridos. Fornecem-nos dez cubos, cada um etiquetado com uma mostra de cor. Há um cubo para o ciano, outro para o amarelo, mais outro para o vermelho, e assim sucessivamente. Depois entregam-nos pedaços de papel coloridos, um por um, e nos dizem para os classificarmos pela cor. Cada papel que recebemos é ligeiramente diferente. Como existe um número infinito de cores no mundo, nunca teremos dois pedaços de papel com a mesma cor exata. Às vezes é fácil indicar em que cubo deve ser colocado o papel colorido, mas em outras acaba sendo difícil. Um papel que é meio vermelho e laranja poderia ir em qualquer um desses cubos, mas temos que lhe atribuir um, vermelho ou laranja, ainda que a seleção acabe sendo aleatória. (O importante deste exercício é mostrar que o cérebro deve classificar padrões. As regiões do córtex fazem isto, mas não há nada equivalente a um cubo para colocar os padrões dentro.)

Agora nos outorgam a tarefa adicional de buscar sequências. Damo-nos conta de que vermelho-vermelho-ciano-roxo-laranja-ciano aparece com frequência. Podemos chamá-la de sequência "vvcrlc". Temos de advertir que seria impossível reconhecer uma sequência se primeiro não tivéssemos classificado os padrões em tipos diferentes. Sem ter classificado antes cada pedaço de papel em uma das dez categorias, não poderíamos afirmar que as duas sequências são iguais.

Portanto, agora estamos preparados para agir. Vamos olhar todos os padrões de entrada —os pedaços coloridos de papel que entram desde as regiões corticais

inferiores—; vamos classificá-los e depois buscar sequências. Ambos os passos, classificação e formação de sequências, são necessários para criar representações invariáveis, e cada uma das regiões corticais os efetuam.

O processo de formar sequências dá resultado quando uma entrada é ambígua, como um pedaço de papel que se encontra entre o vermelho e o laranja. Temos que lhe escolher um cubo ainda que não estejamos seguros se é bem mais vermelho ou bem mais laranja. Se conhecemos a sequência mais provável para esta série de entradas, utilizaremos o dito conhecimento para decidir como classificar a entrada ambígua. Se achamos que estamos na sequência “vvcrlic” porque acabamos de obter dois vermelhos, um ciano e um roxo, esperamos que o próximo papel seja laranja. Porém chega esse próximo papel e o mesmo não é laranja, mas de uma cor rara entre o vermelho e o laranja. Talvez seja inclusive um pouco mais vermelho que o laranja. Porém conhecemos a sequência “vvcrlic” e esperamos ela, de modo que colocamos o papel no cubo laranja. Utilizamos o contexto das sequências conhecidas para resolver a ambiguidade.

Vemos que este fenômeno acontece continuamente em nossas experiências cotidianas. Quando a gente fala, nossas palavras em particular com frequência não podem se entender fora do contexto. No entanto, quando escutamos uma palavra ambígua dentro de uma oração, essa ambiguidade não nos deixa travados. Entendemos ela. De forma similar, as palavras escritas a mão às vezes acabam sendo ininteligíveis fora do contexto, mas são legíveis dentro de uma oração escrita completa. Na maioria das vezes não nos damos conta de que estamos completando informação ambígua ou incompleta das nossas memórias de sequências. Escutamos o que esperamos escutar e vemos o que esperamos ver, pelo menos quando o que escutamos e vemos se encaixa com a experiência passada.

Note que a memória das sequências não só nos permite resolver a ambiguidade da entrada presente, mas também prever que entrada deve aparecer a seguir. Enquanto seu eu cortical classifica papéis de cores, você pode dizer à pessoa “entrada” que lhe passa os papéis: “Ouça, se você está com alguma dificuldade para decidir o que me passar a seguir, segundo a minha memória, deve ser um laranja”. Ao reconhecer uma sequência de padrões, uma região cortical predirá seu próximo padrão de entrada e indicará à região inferior que esperar.

Uma região do córtex não só aprende sequências conhecidas, mas também como modificar suas classificações. Digamos que começamos com um conjunto de cubos etiquetados “ciano”, “amarelo”, “vermelho”, “roxo” e “laranja”. Estamos preparados para reconhecer a sequência “vvcrlic”, bem como outras combinações destas cores. Mas o que acontece se uma cor efetua uma mudança importante? O que acontece se cada vez que vemos a sequência “vvcrlic” o roxo está um pouco distorcido? A nova cor parece-se mais ao anil. De modo que mudamos o cubo roxo para que

seja o cubo "anil". Agora os cubos encaixam melhor com o que vemos; reduzimos a ambiguidade. O córtex é flexível.

Nas regiões corticais, as classificações de abaixo-acima e as sequências de acima-abaixo interagem de forma constante, mudando por completo nossas vidas. Esta é a essência da aprendizagem. De fato, todas as regiões do córtex são "plásticas", o que significa que podem ser modificadas pela experiência. Recordamos o mundo formando novas classificações e sequências.

Por último, observemos como estas classificações e predições interagem com a próxima região superior. Outra parte da nossa tarefa cortical é transmitir o nome da sequência que vemos ao próximo nível superior, de modo que passamos um pedaço de papel com as letras "vvrlc", que significam pouco em si mesmas para essa região superior; o nome não é mais do que um padrão para ser combinado com outras entradas, classificado e depois colocado em uma sequência de ordem superior. Assim como você, ele segue com atenção as sequências que ele vê. Em um determinado momento, ele poderia dizer-nos: "Ouça, no caso de você custar a decidir o que me passar a seguir, segundo minha memória, predigo que deve ser a sequência 'aavca'". Trata-se, em essência, de uma instrução sobre o que devemos buscar em nosso fluxo de entradas. Nos esforçaremos ao máximo para interpretar o que vemos nesta sequência.

Como muita gente tem escutado o termo classificação de padrões empregado na inteligência artificial e nas pesquisas sobre visão das máquinas, observaremos como difere este processo do que faz o córtex. Para tentar conseguir que as máquinas reconheçam objetos, os pesquisadores costumam criar um molde — digamos a imagem de uma caneca, ou de alguma forma, o protótipo de uma caneca— e depois instruem a máquina para que combine suas entradas com a caneca protótipo. Se é descoberta uma correspondência estreita, o computador indicará que encontrou uma caneca. Mas nosso cérebro não tem moldes semelhantes e os padrões que recebe cada região cortical como entradas não são como fotos. Não recordamos fotos instantâneas do que nossa retina vê, ou fotos instantâneas dos padrões da nossa cóclea ou da nossa pele. A hierarquia do córtex assegura que as memórias dos objetos se distribuem por toda ela; não estão localizadas em um único ponto. Além disso, como cada região da hierarquia forma memórias invariáveis, o que uma região normal do córtex aprende são sequências de representações invariáveis, que em si mesmas são sequências de memórias invariáveis. Não encontraremos uma foto de uma caneca ou de qualquer outro objeto armazenado em nosso cérebro.

Diferente da memória de uma câmera, nosso cérebro recorda o mundo tal como é, não como aparece. Quando pensamos sobre o mundo, recordamos sequências de padrões que correspondem a como são os objetos no mundo e a como se comportam, não a como aparecem através de um sentido particular em um determinado momento. As sequências mediante nas quais experimentamos os

objetos do mundo refletem a estrutura invariável do mesmo mundo. A ordem em que experimentamos as partes do mundo está determinado pela estrutura deste. Por exemplo, podemos subir em um avião diretamente caminhando por uma passarela, mas não pelo balcão de bilhetes. A sequência pela qual experimentamos o mundo é a estrutura real deste, e isso é o que o córtex quer recordar.

Não obstante, não esqueçamos que uma representação invariável de qualquer região do córtex pode se converter em uma predição detalhada de como aparecerá em nossos sentidos transmitindo o padrão para abaixo da hierarquia. Da mesma forma, uma representação invariável no córtex motor pode converter-se em ordens motoras detalhadas específicas de uma situação, propagando o padrão para abaixo na hierarquia motora.

Aspecto de uma Região do Córtex

Agora vamos focar nossa atenção em uma região particular do córtex, uma das caixas da figura 5; a figura 6 mostra uma região do córtex com maior detalhe. Minha meta é lhe ensinar como as células de uma região cortical podem aprender e recordar sequências de padrões, que é o elemento mais essencial para formar representações invariáveis e realizar predições. Começaremos com uma descrição da aparência de uma região cortical, e de como ela está conjugada. As regiões corticais variam muito de tamanho, sendo as áreas sensoriais primárias as maiores. V1, por exemplo, é do tamanho aproximado de um passaporte pelo espaço que ocupa na parte posterior do cérebro. Mas, como já expliquei, na realidade ela está composta por muitas regiões menores do tamanho de balas de pequeno calibre. Suponhamos por enquanto que uma área cortical típica é do tamanho de uma moeda pequena.

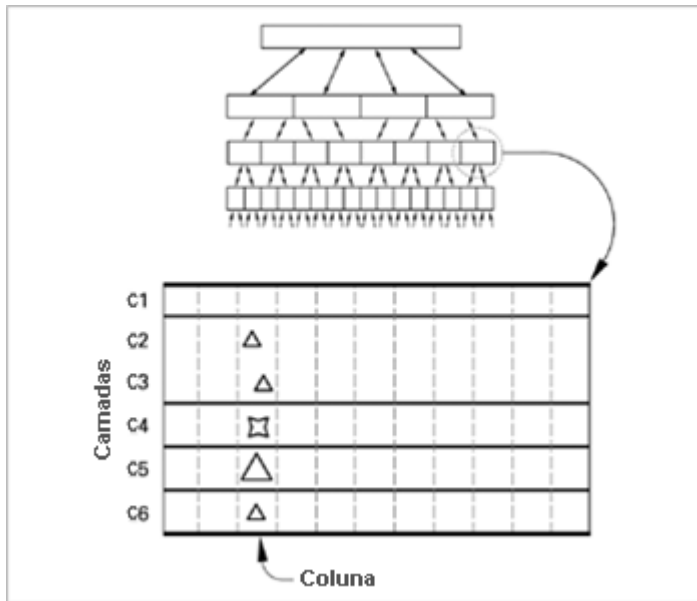


Figura 6. Camadas e colunas em uma região do córtex.

Pensemos nos seis cartões de visita que mencionei no capítulo 3, onde cada um representa uma camada diferente do tecido cortical. Por que dizemos que são camadas? Se tomarmos nossa região cortical do tamanho de uma moeda e colocarmos ela sob um microscópio, veremos que a densidade e tamanho das células varia à medida que nos movemos de acima-abaixo. Estas diferenças definem as camadas. A superior, chamada camada 1, é a mais característica das seis. Tem muito poucas células e consta em essência de um emaranhado de axônios que correm paralelos à superfície cortical. As camadas 2 e 3 apresentam uma aparência similar. Contêm muitos neurônios abarrotados que se chamam células piramidais porque seus corpos celulares se assemelham a pequenas pirâmides. A camada 4 conta com um tipo de células com forma de estrela. A camada 5 tem células piramidais normais, bem como uma classe de células extragrandes em forma de pirâmide. A camada inferior, a 6, também apresenta vários tipos de neurônios únicos. Estas e outras diferenças definem as camadas.

Vemos camadas horizontais, mas com muita frequência os cientistas falam de colunas de células que correm perpendiculares às camadas. Podem-se conceber as colunas como “unidades” verticais de células que funcionam juntas. (O termo coluna suscita muito debate na comunidade da neurociência. Discute-se seu tamanho, função e importância. No entanto, para nosso objetivo podemos concebê-las em termos gerais como uma arquitetura colunar, pois todo mundo está de acordo que ela existe.) As camadas dentro de cada coluna estão ligadas pelos axônios que correm para cima e para abaixo efetuando sinapses. As colunas não aparecem como pequenos pilares nítidos com limites claros —nada no córtex é tão simples—, mas cabe deduzir sua existência por várias linhas de evidência.

Uma razão é que as células alinhadas em vertical de cada coluna tendem a se ativar com os mesmos estímulos. Se observamos detidamente as colunas de V1, descobrimos que algumas respondem aos segmentos de linha que se inclinam em uma direção (/), e outras, aos segmentos de linha que se inclinam em outra direção (\). As células dentro de cada coluna estão muito ligadas, motivo pelo qual a coluna inteira responde aos mesmos estímulos. De forma específica, uma célula ativa da camada 4 faz com que as células superiores das camadas 3 e 2 se ativem, o que depois ocasiona que as células inferiores das camadas 5 e 6 também se ativem. A atividade propaga-se acima e abaixo dentro de uma coluna de células.

Outra razão pela qual falamos de colunas tem que ver com o modo como se forma o córtex. Em um embrião, as células precursoras únicas emigram de uma cavidade cerebral interior até o lugar onde ganha forma o córtex. Cada uma destas células divide-se para criar uns cem neurônios, chamadas microcolunas, que estão ligadas em vertical do modo que acabo de descrever. O termo *coluna* costuma ser usado sem muito rigor para descrever fenômenos diferentes; pode fazer referência à conectividade vertical geral ou a grupos específicos de células do mesmo progenitor. Utilizando a última definição, cabe afirmar que o córtex humano apresenta uma cifra aproximada de várias centenas de milhões de microcolunas.

Para conseguir visualizar esta estrutura colunar, imaginemos que uma microcoluna tem a largura de um cabelo humano. Tomamos milhares de cabelos e os cortamos em segmentos muito reduzidos —digamos da altura de uma *i* minúscula sem o ponto—. Alinhamos todos estes cabelos ou colunas e os colamos lado com lado como se fossem uma escova muito densa. Depois criamos uma lâmina de cabelos longos e extrafinos —que representam os axônios da camada 1— e os colamos horizontalmente na parte superior da lâmina de cabelos curtos. Esta lâmina parecida a uma escova é um modelo simplista da nossa região cortical do tamanho de uma moeda. A informação flui em sua maior parte na direção destes cabelos, horizontalmente na camada 1 e verticalmente nas camadas compreendidas entre a 2 e a 5.

Há mais um detalhe das colunas que é preciso conhecer, e depois passaremos a analisar para que serve. Com uma inspeção minuciosa, vemos que pelo menos 90 por cento das sinapses das células dentro de cada coluna provem de lugares de fora da dita coluna. Algumas conexões chegam de colunas vizinhas; outras, desde pontos equidistantes do outro lado do cérebro. Portanto, como podemos afirmar que as colunas são importantes quando boa parte da nossa conexão cortical se estende lateralmente sobre grandes áreas?

A resposta está no modelo de memória-predição. Em 1979, quando Vernon Mountcastle sustentou que há um único algoritmo cortical, ele também propôs que a coluna cortical é a unidade básica de computação do córtex. No entanto, ele não sabia que funções ela realizava. Acho que a coluna é a unidade básica de predição.

Para que uma coluna prediga quando deve se ativar precisa saber que está passando em outro lugar; daí as conexões sinápticas de um lado a outro.

Em seguida entraremos em mais detalhes, mas esta é uma visão preliminar para compreender por que precisamos desse tipo de conexão no cérebro. Para predizer a próxima nota de uma canção precisamos saber seu nome, em que lugar da canção nos encontramos, quanto tempo tem passado desde a última nota e qual foi essa última nota. O grande número de sinapses que ligam as células de uma coluna com outras partes do cérebro fornece à cada uma o contexto que precisa para predizer sua atividade em muitas situações diferentes.

* * *

A próxima coisa que devemos considerar é como estas regiões corticais do tamanho de uma moeda (e suas colunas) enviam e recebem informação para acima e para abaixo da hierarquia cortical. Primeiro observaremos o fluxo ascendente, que toma uma rota direta, representado na figura 7. Imaginemos que contemplamos uma região cortical com suas milhares de colunas. Nos concentraremos em uma única. As entradas convergentes das regiões inferiores chegam sempre à camada 4, a principal camada de entradas. Ao passar, também formam uma conexão na camada 6 (veremos mais adiante por que é importante). Depois, as células da camada 4 enviam projeções para acima às células das camadas 2 e 3 da sua coluna. Quando uma coluna projeta informação para acima, muitas células das camadas 2 e 3 enviam axônios à camada de entrada da próxima região superior. Deste modo, a informação flui de região em região, ascendendo na hierarquia.

A informação que flui para abaixo da hierarquia cortical toma um caminho menos direto, como é representado na figura 8. As células da camada 6 são as células de saída que se projetam para abaixo desde uma coluna cortical e à camada 1 nas regiões que se encontram abaixo na hierarquia. Na camada 1, os axônios propagam-se por longas distâncias na região cortical inferior. Deste modo, a informação que flui para abaixo na hierarquia de uma coluna possui o potencial de ativar muitas colunas nas regiões que se encontram abaixo dela. Na camada 1 há muito poucas células, mas as das camadas 2, 3 e 5 têm dendritos na camada 1, de modo que estas células podem estimular-se pelas realimentações que percorrem a camada 1. Os axônios provenientes das células das camadas 2 e 3 formam sinapses na camada 5 quando abandonam o córtex, e pensa-se que estimulam células das camadas 5 e 6. Portanto, cabe afirmar que quando a informação flui para abaixo da hierarquia apresenta uma rota menos direta. Pode ramificar-se em muitas direções diferentes mediante sua propagação sobre a camada 1. A informação de realimentação começa em uma célula da camada 6 na região superior e estende-se pela camada 1 na região inferior. Algumas células das camadas 2, 3 e 5 da região inferior estimulam-se, e algumas destas estimulam às células da camada 6, que se projetam à camada 1 nas regiões inferiores da

hierarquia, e assim sucessivamente. (Se você estudar a figura 8, o processo acaba sendo bem mais fácil de seguir.)

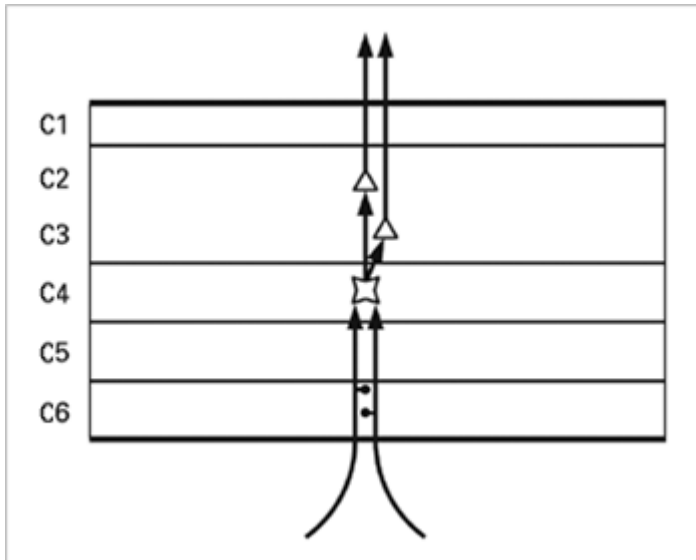


Figura 7. Fluxo ascendente de informação por uma região do córtex.

Tenho aqui uma vista prévia de por que a informação se estende pela camada 1. Converter uma representação invariável em uma predição específica requer a capacidade para decidir momento a momento por qual caminho enviar o sinal quando se propaga para abaixo na hierarquia. A camada 1 proporciona um modo de converter uma representação invariável em outra mais detalhada e específica. Recordemos que podemos lembrar do Discurso de Gettysburg tanto em linguagem falada como escrita. Uma representação comum move-se por um de dois caminhos, um para o pronunciar e outro para o escrever. De forma similar, quando escuto a próxima nota de uma melodia, meu cérebro tem que tomar um intervalo genérico, como uma quinta, e converter na nota específica correta, como dó ou sol. O fluxo horizontal de atividade pela camada 1 proporciona o mecanismo para fazer isto. Para que as predições invariáveis de nível superior se propaguem para abaixo no córtex e se convertam em predições específicas devemos contar com um mecanismo que permita ao fluxo de padrões ramificar-se em cada nível. A camada 1 cumpre os requisitos. Poderíamos predizer sua necessidade inclusive se não soubéssemos que existia.

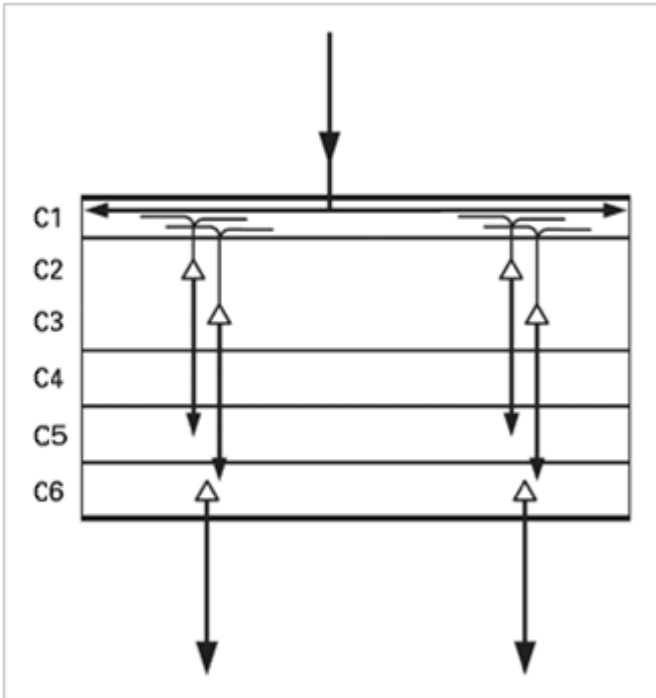


Figura 8. Fluxo descendente de informação por uma região do córtex.

Um dado final sobre anatomia: quando os axônios abandonam a camada 6 para viajar a outros destinos, se encerram em uma substância oleosa branca chamada mielina. Esta assim chamada matéria branca assemelha-se ao isolamento de um cabo elétrico da nossa casa. Ajuda a impedir que os sinais se misturem e as faz viajar mais rápido, a velocidades que superam os 320 quilômetros por hora. Quando os axônios abandonam a matéria branca entram em uma nova coluna cortical da camada 6.

* * *

Para finalizar, consideremos outro método indireto com o que as regiões corticais se comunicam entre si.

Antes de descrever os detalhes, quero que você recorde das memórias auto-associativas analisadas no capítulo 2. Como vimos, as memórias auto-associativas podem se empregar para armazenar sequências de padrões. Quando a saída de um grupo de neurônios artificiais se realimenta para formar a entrada de todos os neurônios e é acrescentado um *delay* à realimentação, os padrões aprendem a se seguir em sequência. Acho que o córtex utiliza o mesmo mecanismo básico para armazenar sequências, ainda que com algumas voltas adicionais. Em vez de formar uma memória auto-associativa com neurônios artificiais, ele forma uma memória auto-associativa com colunas corticais. A saída de todas as colunas é realimentada

à camada 1. Deste modo, a camada 1 contém informação sobre quais colunas estavam ativas na região do córtex.

Repassemos os elementos mostrados na figura 9. Durante muitos anos soube-se que as células particularmente grandes da camada 5 dentro do córtex motor (região M1) estabelecem contato direto com nossos músculos e com as regiões motoras da medula espinhal. Estas células dirigem de forma literal nossos músculos e fazem a gente se mover. Sempre que falamos, digitamos ou realizamos um comportamento complexo, estas células ativam-se e desativam-se de um modo muito coordenado que faz com que nossos músculos se contraíam.

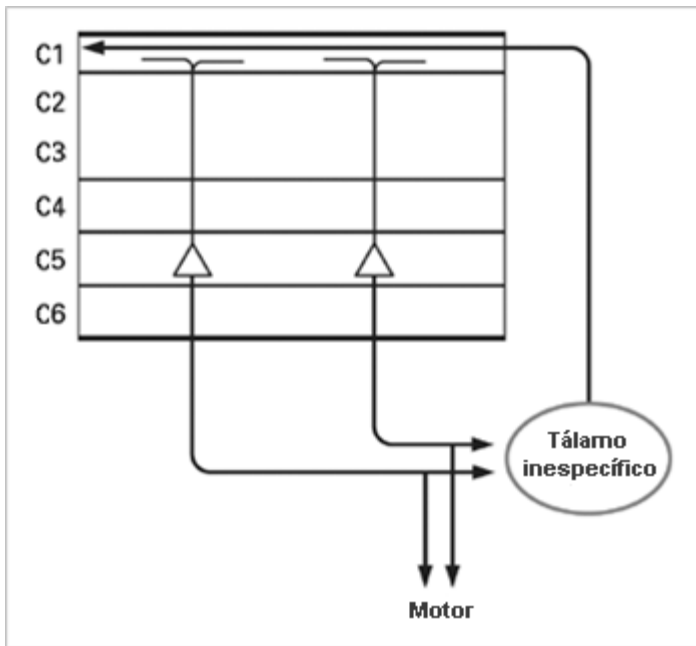


Figura 9. Papel das grandes células da camada 5 no comportamento motor.

Recentemente os pesquisadores descobriram que é provável que as grandes células da camada 5 desempenhem um papel no comportamento de outras partes do córtex, não só nas regiões motoras. Por exemplo, as grandes células da camada 5 do córtex visual projetam-se à parte do cérebro que move os olhos. Portanto, as áreas visuais sensoriais do córtex, como V2 e V4, não só processam entradas visuais, mas também ajudam a determinar o mesmo movimento ocular e, deste modo, o que vemos. As grandes células da camada 5 veem-se ao longo de todo o neocórtex em cada uma das regiões, o que sugere que elas têm um papel mais extenso em todo tipo de movimentos.

Além de gozar de um papel no comportamento, os axônios destas células grandes da camada 5 podem dividir-se em dois. Um ramo vai à parte do cérebro chamada tálamo, representado por um objeto redondo na figura 9. O tálamo humano é do

tamanho e da forma de dois pequenos ovos de pássaro. O mesmo encontra-se no centro do cérebro, no topo do cérebro velho, e está rodeado por matéria branca e pelo córtex. O tálamo recebe muitos axônios de todas as partes do córtex e os reenvia a essas mesmas áreas. Muitos dos detalhes dessas conexões são conhecidos, mas o mesmo tálamo é uma estrutura complexa e seu papel não está bem esclarecido. No entanto, ele é essencial para a vida normal; um tálamo danificado leva a um estado vegetativo persistente.

Há um par de rotas do tálamo ao córtex, mas só uma nos interessa agora. Esta rota começa com as grandes células da camada 5 que se projetam até uma classe de células talâmicas consideradas inespecíficas. As células inespecíficas reprojeta axônios à camada 1 sobre muitas diferentes regiões do córtex. Por exemplo, as células da camada 5, através das regiões V2 e V4, enviam axônios ao tálamo e este devolve a informação às células da camada 1 através de V2 e V4. Outras partes do córtex fazem o mesmo; as células da camada 5 projetam-se através de múltiplas regiões corticais até o tálamo, que reenvia a informação à camada 1 dessas mesmas regiões e associadas. Proponho que este circuito é equivalente à realimentação com *delay* que permite aos modelos da memória associativa aprender sequências.

Tenho mencionado duas entradas à camada 1. As regiões superiores do córtex propagam a atividade pela camada 1 nas regiões inferiores. As colunas ativas de uma região também estendem a atividade pela camada 1 na mesma região através do tálamo. Podemos ilustrar estas entradas à camada 1 como o nome de uma canção (entrada desde acima) e onde nos encontramos em uma canção (atividade com *delay* das colunas ativas da mesma região). Deste modo, a camada 1 transporta muita da informação que precisamos para predizer quando deve se ativar uma coluna —o nome da sequência e onde nos achamos nela—. Utilizando estes dois sinais da camada 1, uma região do córtex é capaz de aprender e recordar múltiplas sequências de padrões.

Como Funciona uma Região do Córtex: Os Detalhes

Com estes três circuitos em mente —padrões convergentes ascendendo pela hierarquia cortical, padrões divergentes descendo pela hierarquia cortical, e uma realimentação com *delay* através do tálamo— podemos começar a contemplar como uma região do córtex realiza as funções que precisa. O que queremos saber é o seguinte:

1.

Como uma região do córtex classifica suas entradas (como os cubos)?

2.

Como ela aprende sequências de padrões (como os intervalos de uma melodia ou o “olho nariz olho” de um rosto)?

3.

Como ela forma um padrão constante ou “nome” para uma sequência?

4.

Como ela realiza previsões específicas (aguardar o trem à hora precisa ou prever a nota exata de uma melodia)?

Começamos assumindo que as colunas de uma região do córtex se parecem aos cubos que temos empregado para classificar nossas entradas de papéis coloridos. Cada coluna representa a etiqueta de um cubo. As células da camada 4 de cada coluna recebem fibras de entrada de várias regiões abaixo e se ativarão se contarem com a combinação adequada de entradas. Quando uma célula da camada 4 se ativa, ela “vota” que a entrada coincide com sua etiqueta. Assim como na analogia da classificação de papéis, as entradas podem ser ambíguas, de modo que várias colunas poderiam ser correspondências possíveis para a entrada. Queremos que nossa região cortical decida sobre uma interpretação; se o papel é vermelho ou laranja, mas não ambas as coisas. Uma coluna com uma entrada forte deve impedir as demais colunas de se ativarem.

Os cérebros contam com células inibidoras que fazem justamente isso. Inibem intensamente os demais neurônios vizinhos do córtex, permitindo na prática que se tenha um ganhador. Estas células inibidoras só afetam a área que rodeia uma coluna. Portanto, ainda que se tenha uma grande inibição, muitas colunas de uma região podem continuar ativas ao mesmo tempo. (Nos cérebros reais, nada é representado com um único neurônio ou com uma única coluna.) Para facilitar seu entendimento, imaginemos que uma região escolha uma coluna ganhadora. Mas recordemos no fundo da nossa mente que é provável que sejam ativadas muitas colunas ao mesmo tempo. O processo real que emprega uma região cortical para classificar entradas e como aprende a fazer é complexo e não se entende muito bem. Para não eternizar-nos com estes temas, presumiremos que nossa região do córtex tem classificado suas entradas como atividade em um conjunto de colunas. Podemos focar-nos então na formação de sequências e nomes para estas.

Como armazena nossa região cortical a sequência destes padrões classificados? Já sugeri uma resposta a esta pergunta, mas agora aprofundarei em mais detalhes. Imaginemos que somos uma coluna de células e a entrada de uma região inferior faz com que uma das células da nossa camada 4 se ative. Ficamos contentes, e nossa célula da camada 4 provoca que também sejam ativadas células das

camadas 2 e 3, depois as da 5 e logo após as da 6. A coluna inteira ativa-se quando é conduzida desde abaixo. Nossas células das camadas 2, 3 e 5 possuem cada uma milhares de sinapses na camada 1. Se algumas destas sinapses estão ativas quando nossas células das camadas 2, 3 e 5 se ativam, as sinapses se fortalecem. Se isto ocorre com a frequência necessária, as sinapses da camada 1 ganham a força suficiente para fazer com que as células das camadas 2,3 e 5 se ativem inclusive quando não se tenha ativado uma célula da camada 4, o que quer dizer que partes da coluna podem se ativar sem receber uma entrada de uma região inferior do córtex. Deste modo, as células das camadas 2, 3 e 5 aprendem a prever quando devem se ativar se baseando no padrão da camada 1. Uma vez que aprende, a coluna pode se ativar parcialmente mediante a memória. Quando uma coluna se ativa através das sinapses da camada 1, está se antecipando a ser estimulada desde abaixo. É uma predição. Se a coluna pudesse falar, diria: "Quando estive ativa no passado, este conjunto particular de sinapses da minha camada 1 esteve ativo, de modo que quando eu voltar a ver este conjunto particular, me ativarei antecipadamente".

Recordemos que metade da entrada à camada 1 provém das células da camada 5 das colunas e regiões vizinhas do córtex. Esta informação representa o que estava acontecendo momentos antes. Representa colunas que estavam ativas antes que nossa coluna se ativasse. É o intervalo anterior na melodia, ou a última coisa que vi, ou a última coisa que senti, ou o fonema anterior no enunciado que estou escutando. Se a ordem na qual aparecem estes padrões ao longo do tempo é constante, as colunas o aprenderão. Se ativarão uma a uma na sequência apropriada.

Outra metade da entrada à camada 1 provém das células da camada 6 de regiões superiores na hierarquia. Esta informação é mais estacionária. Representa o nome da sequência que estamos experimentando no momento. Se nossas colunas são intervalos musicais, é o nome da melodia. Se nossas colunas são fonemas, é a palavra falada que estamos escutando. Se nossas colunas são palavras faladas, o sinal desde acima é o discurso que estamos pronunciando. Portanto, a informação da camada 1 representa tanto o nome de uma sequência como o último elemento desta. Deste modo, uma coluna particular pode ser compartilhada entre muitas sequências diferentes sem criar confusão. As colunas aprendem a se ativar no contexto e ordem precisos.

Antes de continuar avançando, é preciso assinalar que as sinapses da camada 1 não são as únicas que participam na aprendizagem quando uma coluna deve ser ativada. Como já mencionei, as células recebem e enviam entradas a muitas colunas ao redor. Recordemos que mais de 90 por cento das sinapses provem de células alheias à coluna, e a maioria destas sinapses não estão na camada 1. Por exemplo, as células das camadas 2, 3 e 5 possuem milhares de sinapses na camada 1, mas também milhares em suas próprias camadas. A ideia geral é que

as células querem toda informação que lhes ajude a prever quando serão estimuladas desde abaixo. Por regra geral, a atividade nas colunas próximas apresenta uma forte correlação e, portanto, vemos muitas conexões diretas com colunas vizinhas. Por exemplo, se uma linha move-se por nosso campo de visão, ela ativará sucessivas colunas. No entanto, a informação necessária para prever a atividade de uma coluna é com frequência mais global, e aí é onde desempenham um papel as sinapses da camada 1. Se fôssemos uma célula ou uma coluna, não saberíamos o que significam essas sinapses, mas só que elas nos ajudam a prever quando devemos nos ativar.

* * *

Passemos agora a analisar como uma região do córtex forma um nome para uma sequência aprendida. Novamente, imaginemos que somos uma região do córtex. Nossas colunas ativas mudam à cada nova entrada. Conseguimos aprender a ordem na qual nossas colunas se ativam, o que significa que algumas das células das nossas colunas se ativam antes da chegada de entradas provenientes das regiões inferiores. Que informação enviamos às regiões superiores na hierarquia do córtex? Já vimos que nossas células das camadas 2 e 3 enviam seus axônios às próximas regiões superiores. A atividade destas células é a entrada das regiões superiores. Mas isso é um problema. Para que a hierarquia funcione temos que transmitir um padrão constante durante as sequências aprendidas; temos de passar o nome de uma sequência, não os detalhes. Antes de aprender uma sequência podemos passar os detalhes, mas após tê-la aprendido e sermos capazes de prever que colunas se ativarão só deveremos transmitir um padrão constante. No entanto, quando ela está em curso, ainda não criamos um nome para a dita sequência. Passaremos todo padrão mutável sem ter em conta se podemos o prever ou não. Quando todas as colunas se ativam, suas células das camadas 2 e 3 enviarão um novo sinal que ascenderá pela hierarquia. O córtex precisa de um modo de manter constante a entrada à próxima região durante as sequências aprendidas. Precisamos de um modo de desativar a entrada das células das camadas 2 e 3 quando uma coluna prever sua atividade ou, de forma alternativa, manter as células ativas quando a coluna não é capaz de prever. Este é o único modo de criar um padrão constante.

quando uma coluna aprende a predizer sua atividade, a célula da camada 3b permanece quieta. Juntas, as células da camada 2 e 3b cumprem nosso requisito. Antes de aprender, ambas células se ativam e desativam com a coluna, mas após o treinamento a célula da camada 2 está sempre ativa, e a da camada 3b, quieta.

Como estas células aprendem a fazer isto? Consideremos primeiro como desativar a célula da camada 3b quando sua coluna consegue predizer sua atividade. Digamos que há outra célula colocada justo acima da célula da camada 3b na camada 3a. Esta célula também tem dendritos na camada 1. Sua única tarefa consiste em impedir que a célula da camada 3b se ative quando ver o padrão apropriado na camada 1. Quando a célula da camada 3a vê o padrão aprendido na camada 1, ativa de imediato uma célula inibidora que impede estimular a célula da camada 3b. É tudo o que se precisaria para deter a ativação da célula da camada 3b quando a coluna consegue predizer sua atividade.

Analisemos agora a tarefa mais difícil de manter ativa a célula da camada 2 durante uma sequência conhecida de padrões. É mais difícil porque um conjunto diverso de células da camada 2 em muitas colunas diferentes precisaria permanecer ativo junto, ainda que suas colunas particulares não estivessem ativas. Assim é como acho que poderia ocorrer: as células da camada 2 poderiam aprender a ser estimuladas somente desde as regiões superiores na hierarquia do córtex. Poderiam formar sinapses preferenciais com os axônios procedentes das células da camada 6 das regiões mais altas. Deste modo, as da camada 2 representariam o padrão de nome constante das regiões superiores. Quando uma região mais alta do córtex enviasse um padrão à camada 1 da região de abaixo, seria ativado um conjunto de células da camada 2 da região inferior, representando a todas as colunas que são membros da sequência. Já que estas células da camada 2 também se reprojam à região superior, formariam um grupo de células semiestável. (Não é provável que estas células permaneçam sempre ativas; é mais provável que se ativem de forma sincrônica, seguindo uma espécie de ritmo.) É como se a região superior enviasse o nome de uma melodia à camada 1 abaixo. Este fato faz com que seja ativado um conjunto de células da camada 2, um para cada uma das colunas que se ativarão quando a melodia for escutada.

A soma de todos estes mecanismos permite ao córtex aprender sequências, efetuar predições e formar representações constantes, ou “nomes” para as sequências. Estas são as operações básicas para se formar representações invariáveis.

* * *

Como efetuamos predições sobre acontecimentos que nunca tínhamos visto antes? Como decidimos entre múltiplas interpretações de uma entrada? Como uma região do córtex realiza predições específicas partindo de memórias invariáveis? Já contribuí com vários exemplos, como a predição da próxima nota exata de uma

melodia quando nossa memória não recorda mais do que o intervalo entre as notas, a parábola do trem e a recitação do Discurso de Gettysburg. Nestes casos, o único modo de resolver o problema é empregar a última informação específica para converter uma predição invariável em uma predição específica. Outro modo de resolver desde o ponto de vista do córtex é afirmar que temos que combinar a informação de alimentação para adiante (a entrada real) com a informação de realimentação (uma predição em uma forma invariável).

Vejam os um simples exemplo de como acho que isso é realizado. Digamos que a nossa região do córtex lhe tem dito que espere o intervalo musical de uma quinta. As colunas da nossa região representam todos os intervalos específicos possíveis, como dó-mi, dó-sol, ré-lá, etc. Precisamos decidir qual das nossas colunas deve se ativar. Quando a região de cima nos indica que devemos esperar uma quinta, isto provoca com que as células da camada 2 se ativem em todas as colunas que são quintas, como dó-sol, ré-la e mi-si. As células da camada 2 das colunas que representam outros intervalos nos quais participa ré, como ré-mi e ré-si, têm uma entrada parcial. Portanto, na camada 2 temos atividade em todas as colunas que são quintas, e na camada 4 temos uma entrada parcial em todas as colunas que representam intervalos que incluem ré. A interseção destes dois conjuntos representa nossa resposta, a coluna que representa o intervalo ré-la (se veja figura 11).

Como encontra o córtex esta interseção? Recordemos que já mencionei o fato de que os axônios das células das camadas 2 e 3 costumam formar sinapses na camada 5 quando abandonam o córtex e, de forma similar, os axônios que se aproximam à camada 4 desde regiões inferiores do córtex efetuam uma sinapse na camada 6. A interseção destas duas sinapses (de acima-abaxo e de abaxo-acima) proporciona-nos o necessário. Uma célula da camada 6 que recebe estas duas entradas ativas se estimulará. Uma célula da camada 6 representa o que uma região do córtex acha que acontece, uma predição específica. Se uma célula da camada 6 pudesse falar, talvez diria: "Sou parte de uma coluna que representa o intervalo musical ré-la. As outras colunas significam outras coisas. Falo por minha região do córtex. Quando me ativo, quer dizer que acho que o intervalo ré-la está ocorrendo ou ocorrerá. Poderia ativar-me porque a entrada de acima-abaxo procedente dos ouvidos provocou com que a célula da camada 4 da minha coluna excitara à coluna inteira. Ou minha atividade talvez signifique que tenhamos reconhecido uma melodia e estamos predizendo este próximo intervalo específico. De qualquer modo, meu trabalho é indicar às regiões inferiores do córtex no que pensamos que acontece. Represento nossa interpretação do mundo, independentemente se é certa ou só imaginada".

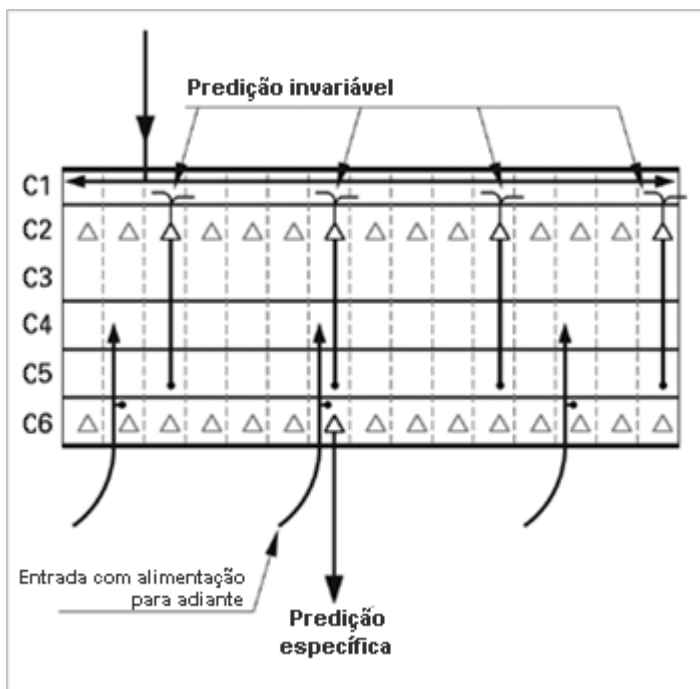


Figura 11. Como uma região do córtex efetua predições específicas partindo de memórias invariáveis.

Vou descrever isto empregando outra imagem mental. Imaginemos dois pedaços de papel com uma porção de buracos perfurados. Os buracos de um papel representam as colunas que têm células ativas da camada 2 ou 3, nossa predição invariável. Os buracos do outro papel representam colunas com uma entrada parcial desde abaixo. Se colocamos um pedaço de papel em cima do outro, alguns dos buracos formarão uma fila e outros não. Os buracos que formam filas representam as colunas que pensamos que devem se ativar.

Este mecanismo não só realiza predições específicas, mas também resolve ambiguidades das entradas sensoriais. Com muita frequência a entrada a uma região do córtex será ambígua, como temos visto com os papéis de cores, ou quando escutamos uma palavra semi-incompreensível. Este mecanismo de correspondência de abaixo-acima/acima-abaixo permite-nos decidir entre duas ou mais interpretações. E uma vez que decidimos, transmitimos nossa interpretação à região abaixo.

Em todo momento da nossa vida consciente, cada região do neocórtex está comparando um conjunto de colunas inesperadas estimuladas desde acima com o conjunto de colunas observadas estimuladas desde abaixo. Quando os dois conjuntos se cruzam surge a nossa percepção. Se tivéssemos uma entrada perfeita desde abaixo e predições perfeitas, o conjunto de colunas percebidas estaria contido sempre no conjunto de colunas preditas. Mas com frequência carecemos de tal harmonia. O método de combinar a predição parcial com a entrada parcial

resolve a entrada ambígua, completa as partes de informação que faltam e decide entre visões alternativas. Assim é como combinamos um intervalo de tom invariável com a última nota escutada para predizer a próxima nota específica em uma melodia. Assim é como decidimos se a foto é de um vaso ou de duas caras. Assim é como dividimos nosso fluxo motor para escrever ou pronunciar o Discurso de Gettysburg.

Por último, além de projetar às regiões corticais inferiores, as células da camada 6 podem reenviar sua saída às células da camada 4 da sua própria coluna. Quando fazem isto, nossas predições se convertem em entrada. É o que fazemos quando sonhamos acordados ou pensamos. Isso nos permite ver as consequências das nossas próprias predições. Fazemos isto muitas horas ao dia quando planejamos o futuro, ensaiamos discursos e nos preocupamos com acontecimentos futuros. Stephen Grossberg, que levou muito tempo modelando o córtex, chama isto de “realimentação dobrada”. Eu prefiro chamar de “imaginação”.

* * *

Um último tema antes de deixar este epígrafe. Já assinalei que na maioria das vezes o que vemos, escutamos ou sentimos depende muito das nossas próprias ações. O que vemos depende de onde efetuam a sacada ocular nossos olhos e como giramos a cabeça. O que sentimos depende de como movemos nossos membros e nossos dedos. O que escutamos depende às vezes do que dizemos e fazemos.

Portanto, para predizer o que veremos a seguir temos que saber que ações estamos fazendo. O comportamento motor e a percepção sensorial são muito interdependentes. Como podemos fazer predições se o que sentimos a seguir é em boa medida o resultado das nossas próprias ações? Por sorte, há uma solução surpreendente e elegante para este problema, ainda que muitos dos detalhes não sejam compreendidos.

A primeira descoberta surpreendente é que a percepção e o comportamento são quase idênticos. Como já mencionei, a maioria —se não todas— as regiões do córtex, inclusive as áreas visuais, participam na criação do movimento. As células da camada 5 que se projetam ao tálamo e depois à camada 1 parecem ter também uma função motora porque se projetam de forma simultânea às áreas motoras do cérebro velho. Deste modo, dispõe-se do conhecimento de “o que acaba de acontecer” —tanto sensorial como motor— na camada 1.

O segundo fato surpreendente, e consequência do primeiro, é que o comportamento motor também deve ser representado em uma hierarquia de representações invariáveis. Geramos os movimentos necessários para levar a cabo uma ação particular, pensando em fazer de uma forma invariável não detalhada. À medida que a ordem motora desloca-se hierarquia abaixo, a mesma é traduzida

em sequências complexas e detalhadas que se requerem para realizar a atividade que esperamos fazer. Isto acontece tanto no córtex “motor” como no “sensorial”, o que torna confusa a distinção entre ambas. Se a região IT do córtex visual percebe “nariz”, o mero ato de mudar a representação de “olho” gerará a sacada ocular necessária para tornar real esta predição. A sacada ocular particular necessária para passar de ver um nariz para ver um olho varia dependendo de onde estiver o rosto. Um rosto próximo requer uma sacada ocular maior; um rosto mais distante, uma sacada ocular menor. Um rosto inclinado requer uma sacada ocular em um ângulo diferente de outra para um rosto horizontal. Os detalhes da sacada ocular precisa determinam-se enquanto a predição de ver o “olho” avança para V1. A sacada ocular torna-se cada vez mais específica quanto mais desce, dando como resultado a que acaba nas fóveas justo na mosca, ou bem perto.

Analisemos outro exemplo. Para que eu me desloque fisicamente da sala à cozinha, tudo o que meu cérebro tem que fazer é mudar da representação invariável da sala para a representação invariável da cozinha. Esta mudança provoca um desdobre complexo de sequências. O processo de gerar a sequência de predições do que verei, sentirei e escutarei enquanto caminhar da sala à cozinha também gera a sequência das ordens motoras que me farão caminhar da sala à cozinha e mover os olhos enquanto efetuar isto. A predição e o comportamento motor funcionam de mãos dadas, enquanto os padrões fluem para acima e para abaixo da hierarquia cortical. Por estranho que pareça, quando nosso próprio comportamento está incluído, nossas predições não só precedem à sensação, mas a determinam. Pensar em passar ao próximo padrão de uma sequência provoca uma predição em cascata do que devemos experimentar a seguir. À medida que se desdobra a predição em cascata, são geradas as ordens motoras necessárias para cumpri-la. Pensar, predizer e agir fazem parte do mesmo desdobre de sequências que descem pela hierarquia cortical.

“Fazer” pensando, o desdobre paralelo da percepção e o comportamento motor, constitui a essência do que se chama comportamento orientado a uma meta. O dito comportamento é o santo graal da robótica, e está incorporada no tecido do córtex.

Podemos desligar nosso comportamento motor, certamente. Posso pensar em ver algo sem o ver na realidade e posso pensar em ir à cozinha sem o fazer na realidade. Mas pensar em fazer algo é, de forma literal, o começo do nosso modo de o fazer.

Fluxos Ascendentes e Descendentes

Retrocedamos um pouco para pensar um pouco mais sobre como se move a informação para acima e para abaixo da hierarquia cortical. Quando nos deslocamos pelo mundo fluem entradas mutáveis às regiões inferiores do córtex. Cada região tenta interpretar sua corrente de entradas como parte de uma

sequência de padrões conhecida. As colunas tentam prever sua atividade. Se conseguem isto, passarão um padrão estável, o nome da sequência, à próxima região superior. Novamente, é como se a região dissesse: "Estou escutando uma canção; este é seu nome. Posso manejar os detalhes".

Mas o que ocorre se chega um padrão inesperado, uma nota imprevista? Ou o que ocorre se vemos algo que não pertence a um rosto? O padrão inesperado passa de forma automática à próxima região cortical superior. Ocorre de modo natural quando se estimulam as células da camada 3b que não faziam parte da sequência esperada. A região superior pode ser capaz de compreender este novo padrão como a próxima parte da sua sequência própria. Poderia dizer: "Vejo que chegou uma nova nota. Talvez é a primeira da próxima canção do álbum. Parece isto, de modo que predigo que já passamos para a próxima canção. Região inferior, aqui está o nome da próxima canção que acho que você deverá escutar". Mas se não ocorre este reconhecimento, o padrão inesperado continuará se propagando para acima da hierarquia cortical até que alguma região superior possa o interpretar como parte da sua sequência de fatos normal. Quanto mais precise ascender o padrão inesperado, mais regiões do córtex participam na resolução da entrada inesperada. Por último, quando uma região elevada da hierarquia pensa que pode compreender o fato inesperado, é gerada uma nova predição. Esta se propaga para abaixo na hierarquia até onde pode avançar. Se a nova predição não estiver correta, o erro será detectado e novamente ascenderá pela hierarquia até que alguma região seja capaz de o interpretar como parte da sua sequência ativa em curso. Deste modo, vemos que os padrões observados fluem para acima na hierarquia, e as predições, para abaixo. Idealmente, em um mundo conhecido e predizível, a maior parte do fluxo de padrões ascendente e descendente ocorre com rapidez e nas regiões inferiores do córtex. O cérebro trata em seguida de encontrar uma parte do seu modelo do mundo que seja compatível com qualquer entrada inesperada. Só então pode saber razoavelmente o que esperar depois.

Se caminho por um cômodo conhecido da minha casa serão propagados poucos erros para acima do meu córtex. As sequências bem aprendidas da minha casa podem ser manejadas nas seções inferiores da hierarquia visual e motora. Conheço tão bem o cômodo que inclusive posso o percorrer na escuridão. Minha familiaridade com o ambiente liberta na prática a maior parte do meu córtex para outras tarefas, como pensar em cérebros e escrever livros. No entanto, se eu estivesse em um cômodo desconhecido, sobretudo um que fosse diferente de todos os que já tinha visto antes, não só precisaria olhar para ver por onde eu avançava, mas padrões inesperados ascenderiam cada vez mais acima da hierarquia cortical. Quanto menos concordar minha experiência sensorial com as sequências aprendidas, mais erros surgirão. Nesta situação nova, já não posso pensar em cérebros porque a maior parte do meu córtex está ocupado com o problema de percorrer o cômodo. É uma experiência comum para as pessoas que saem de um avião em um país estrangeiro. Ainda que as estradas talvez pareçam

semelhantes às que estamos acostumados, é provável que os carros circulem pela mão contrária, o dinheiro seja estranho, a língua acabe sendo incompreensível e aprender a encontrar um banheiro pode necessitar de toda a potência do nosso córtex. Não tente ensaiar um discurso enquanto caminha por solo estrangeiro.

A sensação de entendimento repentino, o momento de "ah!", pode ser entendido neste modelo. Imaginemos que estamos olhando um quadro ambíguo. Cheio de manchas de tinta e linhas dispersas, ele não se parece a nada. Carece de sentido. A confusão surge quando o córtex não consegue encontrar nenhuma memória que se corresponda com a entrada. Nossos olhos esquadrinham cada lugar do quadro. Novas entradas percorrem todo o caminho ascendente da hierarquia cortical. As regiões superiores tentam muitas hipóteses diferentes, mas quando as ditas predições percorrem o caminho descendente da hierarquia, todas e cada uma acabam sendo incompatíveis com a entrada, e o córtex se vê obrigado a realizar uma nova tentativa. Durante este tempo de confusão nosso cérebro está totalmente ocupado em entender o quadro. Por fim efetuamos uma predição de alto nível que é a correta. Quando acontece isto, a predição começa no topo da hierarquia cortical e consegue propagar até a base. Em menos de um segundo, à cada região é entregue uma sequência que se encaixa com os dados. Não ascendem mais erros ao topo. Compreendemos que o quadro é entendível; vemos um cão dálmata entre os pontos e manchas (veja a figura 12).



Figura 12. Vê o dálmata?

Pode a Realimentação Conseguir Isso?

Temos sabido durante décadas que as conexões da hierarquia cortical são recíprocas. Se uma região A projeta-se à região B, B projeta-se à região A.

Costuma-se ter mais fibras de axônios indo para adiante do que para atrás. Mas ainda que esta descrição goze de uma ampla aceitação, o paradigma predominante é que a realimentação desempenha um papel menor ou “modulador” no cérebro. A ideia de que um sinal de realimentação possa provocar de forma imediata e precisa que um conjunto diverso de células da camada 2 se estimule não é a postura dominante entre os neurocientistas.

Por que deveria ser assim? Parte da razão, como já mencionei, é que não existe uma necessidade real de se preocupar pela realimentação se não se aceita o papel central da predição. Se pensa-se que a informação flui sem interrupções pelo sistema motor, por que se precisa de realimentação? Outra razão para ignorá-la é que o sinal de realimentação se estende por grandes áreas da camada 1. Por intuição se esperaria que um sinal que se dispersa por uma grande área tenha um efeito menor em muitos neurônios, e de fato, o cérebro conta com vários sinais moduladores semelhantes que não atuam sobre neurônios específicos, mas que mudam atributos globais como o estado de alerta.

A razão final para passar por alto a realimentação baseia-se em como creem muitos cientistas de como funcionam os neurônios. Os neurônios típicos contam com milhares ou centenas de milhares de sinapses. Algumas se localizam longe do corpo celular; outras, justo em cima, bem perto. As sinapses próximas ao corpo celular têm uma grande influência na estimulação da célula. Uma dúzia aproximada de sinapses ativas próximas do corpo celular podem provocar com que seja gerado um pico ou impulso de descarga elétrica. Isto é sabido. No entanto, a vasta maioria das sinapses não estão próximas do corpo da célula. Elas estendem-se pela longa e larga estrutura ramificada dos dendritos. Já que estas sinapses estão muito afastadas do corpo celular, os cientistas tendem a pensar que um pico que chega a uma destas sinapses teria um efeito débil ou quase imperceptível para que o neurônio gere um pico. O efeito de uma sinapse distante já teria se dissipado quando chegasse ao corpo celular.

Como regra geral, a informação que flui para acima da hierarquia cortical se transfere através das sinapses próximas aos corpos celulares. Deste modo, é mais seguro que a informação que ascende pela hierarquia passe de uma região a outra. Mesmo assim, como regra geral, a realimentação que desce pela hierarquia cortical o faz através das sinapses afastadas do corpo celular. As células das camadas 2, 3 e 5 enviam dendritos à camada 1 e formam ali muitas sinapses. A camada 1 é uma massa de sinapses, mas todas estão afastadas dos corpos celulares das camadas 2, 3 e 5. Além disso, uma célula particular da camada 2 só formará algumas sinapses, se acontecer, com uma fibra de realimentação particular. Portanto, alguns cientistas podem se opor à ideia de que um padrão breve da camada 1 possa provocar com que um conjunto de células se estimule nas camadas 2, 3 e 5. Mas isto é precisamente o que a teoria que tenho exposto requer.

A resolução deste dilema é que os neurônios se comportam de forma diferente do modo como o fazem no modelo clássico. De fato, nos anos recentes, um grupo crescente de cientistas tem proposto que as sinapses em dendritos distantes e finos podem desempenhar um papel ativo muito específico na ativação celular. Nestes modelos, essas sinapses distantes comportam-se de forma diferente das sinapses dos dendritos mais grossos próximos ao corpo celular. Por exemplo, se tivesse duas sinapses bem perto uma da outra em um dendrito fino, elas agiriam como "detectores de coincidência". Isto é, se ambas sinapses recebessem um pico de entrada dentro de um quadro de tempo reduzido, poderiam exercer um grande efeito sobre a célula ainda que estivessem afastadas do corpo celular. Poderiam provocar com que o corpo celular gerasse um pico. Como se comportam os dendritos de um neurônio continua sendo um mistério, de modo que não posso dizer muita coisa a respeito. O importante é que o modelo de memória-predição do córtex requer que as sinapses afastadas do corpo celular sejam capazes de detectar padrões específicos.

Pensando a posteriori, parece quase óbvio afirmar que a maioria das milhares de sinapses de um neurônio se limitam a desempenhar um papel modulador. A realimentação em massa e o número exorbitante de sinapses existem por uma razão. Seguindo esta percepção, cabe afirmar que um neurônio típico tem a capacidade de aprender centenas de coincidências precisas nas fibras de realimentação quando estabelecem sinapses sobre dendritos finos. Isso significa que cada coluna do nosso neocórtex é muito flexível desde a perspectiva dos padrões de realimentação que provocam com que ela se ative. Significa que qualquer característica particular pode se associar de maneira precisa com milhares de objetos e sequências diferentes. Meu modelo requer que a realimentação seja rápida e precisa. As células precisam estimular-se quando veem um número de coincidências precisas em seus dendritos distantes. Estes novos modelos de neurônios permitem isto.

Como Aprende o Córtex

Todas as células de todas as camadas do córtex contam com sinapses, e a maioria destas podem se modificar mediante a experiência. Cabe afirmar que a aprendizagem e a memória ocorrem em todas as camadas, em todas as colunas e em todas as regiões do córtex.

Já falei do livro da aprendizagem hebbiana, batizado com esse nome pelo neuropsicólogo canadense Donald O. Hebb. Sua essência é muito simples: quando dois neurônios se ativam ao mesmo tempo, as sinapses entre elas se fortalecem. Agora sabemos que Hebb estava basicamente certo. Certamente, nada na Natureza é tão simples, e nos cérebros reais os detalhes são mais complexos. Nossos sistemas nervosos praticam muitas variações da regra de aprendizagem hebbiana; por exemplo, algumas sinapses mudam sua força em resposta a pequenas variações na sincronização dos sinais neurais; algumas mudanças

sinápticas são de curta duração, e outras, de longa. Mas Hebb limitou-se a estabelecer um modelo para o estudo da aprendizagem, não era uma teoria final, e o dito modelo tem se tornado incrivelmente útil.

Os princípios da aprendizagem hebbiana podem explicar a maioria do comportamento cortical que tenho mencionado neste capítulo. Recordemos que já se demonstrou na década de 1970 que as memórias auto-associativas, empregando o algoritmo hebbiano clássico, são capazes de aprender padrões espaciais e sequências de padrões. O problema principal era que as memórias não conseguiam manejar bem a variação. Segundo a teoria proposta neste livro, o córtex ficou em torno desta limitação em parte acumulando memórias associativas em uma hierarquia e em parte usando uma complexa arquitetura colunar. Este capítulo consagrou-se sobretudo a analisar a hierarquia e seu funcionamento porque é ela que torna poderoso o córtex. Portanto, em vez de dedicar-me a descrever com detalhes minuciosos como poderia aprender cada célula isto ou aquilo, desejo abordar alguns princípios amplos de aprendizagem em uma hierarquia.

Quando nascemos, nosso córtex não sabe nada. Não conhece nossa língua, nossa cultura, nossa casa, nossa cidade, canções, as pessoas com as quais cresceremos, nada. Toda esta informação, a estrutura do mundo, tem que ser aprendida. Os dois componentes básicos da aprendizagem são a classificação de padrões e a construção de sequências. Estes dois componentes complementares da memória interagem. Quando uma região aprende sequências, as entradas das células da camada 4 mudam. Portanto, estas células da camada 4 aprendem a formar novas classificações, que mudam o padrão reprojeto à camada 1, que afeta às sequências.

O básico de formar sequências é agrupar padrões que são parte do mesmo objeto. Um modo de fazer isto é agrupando os padrões que aparecem seguidos no tempo. Se uma menina sustenta um brinquedo na mão e move-o lentamente, seu cérebro pode assumir sem dúvida alguma que a imagem da sua retina é do mesmo objeto um momento após outro e, portanto, o conjunto mutável de padrões pode ser agrupado junto. Em outros momentos são necessárias instruções externas para ajudar a decidir que padrões devem ir juntos. Aprender que as maçãs e as bananas são frutas, mas as cenouras e o aipo não, requer um professor que oriente a agrupar estes artigos como frutas. De qualquer modo, nosso cérebro constrói lentamente sequências de padrões que são similares. Mas quando uma região do córtex constrói padrões, a entrada na próxima região muda. A entrada deixa de representar em sua maioria padrões particulares e passa a representar grupos de padrões. A entrada de uma região muda de notas a melodias, de letras a palavras, de narizes a rostos, e assim sucessivamente. Como as entradas de abaixo-acima de uma região se tornam mais "orientadas à objeto", a região superior do córtex pode agora aprender sequências desses objetos de ordem

superior. Onde antes uma região construía sequências de letras, agora constrói sequências de palavras. O resultado inesperado deste processo é que durante a aprendizagem repetitiva as representações de objetos descem pela hierarquia cortical. Durante os primeiros anos da nossa vida, nossas memórias do mundo formam-se primeiro nas regiões superiores do córtex, mas à medida que aprendemos são re-formadas em partes cada vez mais baixas da hierarquia cortical. Não é que o cérebro as mova; ele tem que reaprendê-las uma e outra vez. (Não sugiro que todas as memórias comecem no topo do córtex. A formação real das memórias é mais complexa. Acho que a classificação de padrões da camada 4 começa abaixo e vai ascendendo. Mas à medida que é feito isto, começamos a formar sequências que depois descem. É a memória das sequências que sugiro que se re-forma cada vez mais abaixo do córtex.) À medida que as representações simples vão descendo, as regiões mais elevadas são capazes de aprender padrões mais complexos e sutis.

Pode-se comprovar a criação e movimento descendente da memória hierárquica observando como aprende uma criança. Analisemos como aprendemos a ler. A primeira coisa que aprendemos é a reconhecer letras impressas individuais. É uma tarefa lenta e difícil que requer um esforço consciente. Depois passamos a reconhecer palavras simples. Novamente, acaba sendo difícil e lento no princípio, inclusive para palavras de três letras. A criança pode ler cada letra em sucessão e pronunciar as letras uma a uma, mas precisa de muita prática antes de ser capaz de reconhecer a palavra em si como um todo. Depois de aprender palavras simples, lutamos com palavras multisilábicas complexas. A princípio pronunciamos cada sílaba concatenando-as como fizemos com as letras quando aprendemos palavras simples. Após anos de prática, uma pessoa pode ler rapidamente. Chegamos a um ponto no qual não vemos todas as letras individuais, mas reconhecemos palavras inteiras e com frequência frases inteiras em só uma olhada. Não é só porque estamos mais rápidos; estamos reconhecendo as palavras e frases como entidades. Quando lemos uma palavra inteira de uma só vez, seguimos vendo as letras? Sim e não. É evidente que a retina vê as letras e, portanto, também o fazem as regiões de V1. Mas seu reconhecimento ocorre bastante abaixo na hierarquia cortical. Digamos em V2 ou V4. Quando o sinal chega a IT, as letras individuais já não estão representadas. O que a princípio custava o esforço do córtex visual inteiro —reconhecer letras individuais—, agora ocorre mais próximo da entrada visual. À medida que a memória de objetos simples como letras desce pela hierarquia, as regiões superiores têm a capacidade de aprender objetos complexos como palavras e frases.

Aprender a ler música é outro exemplo. A princípio deve se concentrar em cada nota. Com a prática, começa-se a reconhecer sequências de notas comuns, depois frases inteiras. Depois de muita prática, é como se não se visse a maioria das notas. A música da partitura está lá só para recordar a estrutura principal da peça;

as sequências detalhadas foram memorizadas mais abaixo. Este tipo de aprendizagem ocorre tanto nas áreas motoras como nas sensoriais.

Um cérebro jovem é mais lento para reconhecer entradas e efetuar ordens motoras porque as memórias usadas nestas tarefas encontram-se mais acima da hierarquia cortical. A informação tem que fluir o caminho completo acima e abaixo, talvez com múltiplos passes, para resolver os conflitos. Leva tempo para que os sinais neurais percorram acima e abaixo a hierarquia cortical. Um cérebro jovem também não formou ainda sequências complexas no topo e, portanto, não pode reconhecer e reenviar padrões complexos. Não é capaz de entender a estrutura de ordem superior do mundo. Comparado com o de um adulto, a linguagem da criança é simples, sua música é simples e suas interações sociais são simples.

Se estudamos um conjunto particular de objetos uma e outra vez, nosso córtex re-forma as representações da memória desses objetos hierarquia abaixo, o que liberta o topo para aprender relações mais sutis e complexas. Segundo a teoria, isto é o que faz um expert.

Em meu trabalho de projeto de computadores, algumas pessoas surpreendem-se da rapidez com que posso olhar um produto e ver os problemas inerentes em seu projeto. Depois de vinte e cinco anos projetando computadores, conto com um modelo superior à média dos temas associados com os aparelhos de informática portáteis. De modo similar, um pai experiente reconhece com facilidade por que seu filho está aborrecido, enquanto que a outro inexperiente talvez lhe custe manejar a situação. Um diretor empresarial pode ver com facilidade as falhas e as vantagens da estrutura de uma organização, enquanto que o diretor inexperiente não chega a compreender essas coisas. Eles possuem a mesma entrada, mas o modelo do novato não é tão sofisticado. Em todos estes casos e em mais outros milhares começamos aprendendo o básico, a estrutura mais simples. Com o tempo nosso conhecimento desce pela hierarquia cortical e, portanto, no topo temos a oportunidade de aprender estruturas de ordem superior. É esta estrutura de ordem superior que nos outorga experiência. Os experts e os gênios possuem cérebros que veem a estrutura da estrutura e os padrões dos padrões para além do que vê os demais. Podemos nos converter em experts mediante a prática, mas sem dúvida também existe um componente genético.

O Hipocampo: no Topo de Tudo

Três grandes estruturas cerebrais encontram-se sob a lâmina neocortical e comunicam-se com ela. São os núcleos basais, o cerebelo e o hipocampo. Os três existiam antes do neocórtex. Em linhas muito gerais, pode-se dizer que os núcleos basais eram o sistema motor primitivo, o cerebelo aprendia a sincronização precisa das relações de acontecimentos e o hipocampo armazenava memórias de acontecimentos e lugares específicos. Até certo ponto, o neocórtex tem subsumido suas funções originais. Por exemplo, um ser humano nascido sem boa parte do

cerebelo terá deficiências em sincronização e deverá aplicar um esforço mais consciente quando se mover, mas no restante será bastante normal.

Sabemos que o neocórtex é responsável por todas as sequências motoras complexas e pode controlar diretamente nossas extremidades. Não que os núcleos basais careçam de importância, mas o neocórtex tem assumido uma grande parte do controle motor. Devido a isso, tenho descrito sua função geral independente dos núcleos basais e do cerebelo. É provável que alguns cientistas não estejam de acordo com esta proposta, mas é o que tenho usado neste livro e no meu trabalho.

No entanto, o hipocampo é uma outra questão. É uma das áreas do cérebro mais estudadas porque acaba sendo essencial para a formação de novas memórias. Se perdemos ambas as metades do hipocampo (assim como muitas partes do sistema nervoso, ele existe no lado esquerdo e direito do cérebro) ficamos sem capacidade para formar a maioria das novas memórias. Sem o hipocampo somos capazes de continuar falando, caminhando, vendo e escutando, e durante um breve período pareceremos quase normais, mas na realidade estamos profundamente deteriorados: não podemos recordar nada novo. Lembraremos-nos de um amigo que conhecemos antes de perder o hipocampo, mas não conseguiremos recordar uma nova pessoa. Ainda que nos reunissemos com nosso médico cinco vezes ao dia durante um ano, cada vez seria como a primeira. Não teríamos memória dos fatos que ocorreram depois da perda do hipocampo.

Durante muitos anos detestei pensar no hipocampo porque carecia de sentido para mim. Sem dúvida, ele é essencial para a aprendizagem, mas não é o armazém supremo da maior parte do que conhecemos. O neocórtex é. A visão clássica do hipocampo é de que ali se formam memórias novas que mais tarde, ao longo de um período de dias, semanas ou meses, são transferidas ao neocórtex. Isto não fazia sentido para mim. Sabemos que a visão, o som e o tato —nossa corrente de dados sensoriais— fluem diretamente às áreas sensoriais do córtex sem passar primeiro pelo hipocampo. Parecia-me que esta informação sensorial devia formar de modo automático novas memórias no córtex. Por que precisamos de um hipocampo para aprender? Como pode uma estrutura separada como o hipocampo interferir e impedir a aprendizagem no córtex se ele for limitado a transferir a informação de volta ao córtex?

Decidi deixar de lado o hipocampo, imaginando que em algum dia seu papel se clarearia. E este dia chegou no final de 2002, mais ou menos na época em que comecei a escrever este livro. Um dos meus colegas do Redwood Neuroscience Institute, Bruno Olshausen, assinalou que as conexões entre o hipocampo e o neocórtex sugerem que o primeiro é a região suprema do neocórtex e não uma estrutura separada. Segundo esta proposta, o hipocampo ocupa o cume da pirâmide neocortical, o bloco superior da figura 5. O neocórtex apareceu no cenário evolutivo emparedado entre o hipocampo e o resto do cérebro. Ao que

parece, esta proposta do hipocampo como o topo da hierarquia cortical se conhecia há algum tempo, mas eu não estava a par disto. Conversei com vários experts e lhes pedi que me explicassem como esta estrutura em forma de cavalo-marinho podia transferir memórias ao córtex. Ninguém soube como explicar, nem tampouco mencionaram que ele se encontrava no cume da pirâmide cortical, provavelmente por que além de não se encontrar fisicamente na dita posição, ele se liga de forma direta com muitas outras partes mais antigas do cérebro.

Não obstante, percebi instantaneamente que esta nova perspectiva era a solução da minha confusão.

Pensemos sobre a informação que flui dos olhos, ouvidos e pele ao neocórtex. Cada região desta tenta entender o que a dita informação significa. Cada região tenta compreender a entrada em virtude das sequências que conhece. Se ela compreende a entrada, ela diz: "Entendo isto, é parte do objeto que estou vendo. Não passarei os detalhes". Se uma região não compreende a entrada em curso, ela a passa hierarquia acima até que alguma região superior o faça. No entanto, um padrão realmente novo escalará cada vez mais acima da hierarquia. Cada região mais elevada diz: "Não sei o que é isto, não o previ; por que não o confere mais acima?". O efeito líquido é que quando chegamos ao topo da pirâmide cortical o que temos deixado é informação que não se pode compreender com a experiência anterior. Fica-nos a parte da entrada que é realmente nova e inesperada.

Em um dia normal encontramos muitas coisas novas que chegam até o topo —por exemplo, uma história no jornal, o nome da pessoa que conhecemos nesta manhã e o acidente de carro que vimos no caminho de casa—. São estes resíduos inexplicados e inesperados, as coisas novas, que entram no hipocampo e se armazenam ali. Esta informação não será guardará para sempre. Será retransferida ao córtex ou acabará se perdendo.

Dei-me conta de que, à medida que vou ficando mais velho, me custa recordar coisas novas. Por exemplo, meus filhos recordam os detalhes da maioria das obras de teatro que viram no ano passado. Eu não. Talvez seja porque tenho visto tantas em minha vida que raramente encontro algo realmente novo. As novas obras se encaixam nas memórias das obras passadas e a informação não chega a meu hipocampo. Para meus filhos, cada obra é uma novidade e chega ao hipocampo. Se isto for verdade, caberia afirmar que quanto mais sabemos, menos recordamos.

Diferente do neocórtex, o hipocampo possui uma estrutura heterogênea com várias regiões especializadas. Realiza perfeitamente a única tarefa de armazenar em seguida qualquer padrão que vê. Encontra-se na posição perfeita, no topo da pirâmide cortical, para recordar o que é novo. Também está na posição perfeita para recordar essas novas memórias, permitindo que se armazenem na hierarquia cortical, que é um processo um pouco mais lento. No hipocampo podemos recordar instantaneamente um acontecimento novo para ele, mas só recordaremos

algo de forma permanente no córtex se o experimentamos uma e outra vez, seja na realidade ou pensando nisso.

Uma Rota Alternativa para Ascender na Hierarquia

Nosso córtex conta com uma segunda rota importante para passar informação de uma região a outra ascendendo na hierarquia. Esta rota alternativa começa com as células da camada 5 que se projetam no tálamo (uma parte diferente das que analisamos antes) e depois do mesmo às regiões superiores do córtex. Sempre que duas regiões do córtex ligam-se diretamente entre si de forma hierárquica, também o fazem indiretamente através do tálamo. Esta segunda rota só passa informação hierarquia acima, não abaixo. Portanto, quando ascendemos pela hierarquia cortical, existe um caminho direto entre duas regiões e outro indireto através do tálamo.

O segundo caminho apresenta dois modos de operação determinados pelas células do tálamo. Em um modo, a rota está em boa parte fechada, de modo que a informação não flui através dela. No outro modo, a informação flui com precisão entre as regiões. Dois cientistas, Murray Sherman, da Universidade Estadual de Nova York, em Stony Brook, e Ray Guillery, da Escola de Medicina da Universidade de Wisconsin, descreveram esta rota alternativa e postularam que ela poderia ser tão importante como a direta (talvez até mais) que tem constituído o tema deste capítulo até o momento. Tenho uma hipótese sobre o que faz esta segunda rota.

Leia esta palavra: imaginação. A maioria das pessoas pode reconhecê-la com uma única olhada, uma fixação. Agora olhe a letra i no meio da palavra. Agora olhe o ponto sobre a i. Seus olhos podem fixar-se na mesma localização exata, mas em um caso veem a palavra; no próximo, a letra, e no último, o ponto. Se custa-lhe entender isto, tente dizer "ponto", "i" e "imaginação" enquanto olha fixamente o ponto. Em todos os casos entra a mesma informação exata em V1, mas quando chega a uma região superior como IT percebem diferentes coisas, diferentes graus de detalhe. A região IT sabe como reconhecer os três objetos. Pode reconhecer o ponto isolado na letra i e na palavra inteira em uma olhada. Mas quando percebem a palavra inteira, V2, V4 e V1 se ocupam dos detalhes, e tudo o que chega a conhecer IT a respeito é a palavra. Por regra geral, não percebemos as letras individuais enquanto lemos; percebemos palavras ou frases. Mas podemos fazer isto se quisermos. Fazemos esta espécie de mudança de atenção continuamente, mas não costumamos ter consciência disso. Posso estar escutando música de fundo e mal me dar conta da melodia, mas se eu tentar, sou capaz de isolar o cantor ou o baixo. O mesmo som entra em minha cabeça, mas posso focar minhas percepções. A cada vez que coçamos a cabeça, o movimento provoca um forte som interno, mas não costumamos o perceber. No entanto, se concentramo-nos nele, conseguimos o escutar com clareza. Este é outro exemplo da entrada sensorial que pelo geral se maneja em regiões baixas da hierarquia cortical, mas que pode ascender a níveis superiores se lhe prestarmos atenção.

Minha conjectura é que a rota alternativa pelo tálamo é o mecanismo mediante o qual atendemos aos detalhes que normalmente não perceberíamos. Contorna o agrupamento de sequências na camada 2 e envia os dados brutos à próxima região superior do córtex. Os biólogos têm mostrado que a rota alternativa pode se abrir de dois modos. Um é o método que você empregou quando lhe pedi que prestasse atenção a detalhes que não costuma perceber, como o ponto sobre a letra i ou o som ao se coçar a cabeça. O segundo método que pode ativar esta rota é um grande sinal inesperado de abaixo. Se a entrada na rota alternativa tem a força suficiente, ela envia um sinal de alerta à região superior, que pode abrir novamente a rota. Por exemplo, se lhe mostrasse um rosto e lhe perguntasse o que era, você diria: "Rosto". Se lhe mostrasse o mesmo rosto mas tivesse uma estranha cicatriz no nariz, primeiro o reconheceria, mas depois seus níveis inferiores de visão se dariam conta de que algo não vai bem. Este erro obriga a abrir a rota da atenção. Os detalhes tomarão agora o caminho alternativo, evitando o agrupamento que ocorre normalmente, e a cicatriz suscitaria sua atenção. Agora você vê a cicatriz e não só o rosto. Se ela fosse bastante rara, a cicatriz poderia ocupar sua inteira atenção. Deste modo, os acontecimentos pouco habituais suscitam de imediato sua atenção. Este é o motivo pelo qual não podemos evitar de nos fixar nas deformidades e em outros padrões inusitados. Nosso cérebro faz isto de forma automática. No entanto, com frequência os erros não têm a força necessária para abrir a rota alternativa. Por isso, às vezes não nos damos conta quando uma palavra está mau escrita quando a lemos.

Reflexões Finais

Para encontrar e estabelecer um novo modelo científico é necessário buscar os conceitos mais simples capazes de unir e explicar o que eram grandes quantidades de dados distintos. Uma consequência inevitável deste processo é que o pêndulo se incline muito para a simplificação. É provável que se passe por alto detalhes importantes e que dados sejam mal interpretados. Se o modelo se consolida, é inevitável que surjam ajustes que mostrem onde foi muito longe a proposta inicial, onde ficou curta ou onde estava o erro.

Neste capítulo tenho apresentei muitas ideias especulativas sobre o funcionamento do neocórtex. Confio plenamente de que várias das ditas ideias podem acabar sendo errôneas, e é muito provável que todas sejam revisadas. Há, além disso, muitos detalhes que nem sequer mencionei. O cérebro é muito complexo; os neurocientistas que leem este livro saberão que apresentei uma grosseira caracterização das complexidades de um cérebro real. Não obstante, acho que o modelo em seu conjunto é sólido. Não me resta mais do que esperar que as ideias centrais se conservem quando os detalhes mudarem em face de novos dados e entendimento.

Por último, talvez lhe custe aceitar a ideia de que um sistema de memória simples mas grande possa dar como resultado tudo o que fazem os humanos. Seria

possível que você e eu não fôssemos mais do que um sistema de memória hierárquica? Poderiam armazenar-se nossas vidas, credos e ambições em trilhões de sinapses diminutas? Em 1984 comecei a escrever programas de computador de maneira profissional. Tinha escrito pequenos programas antes, mas era a primeira vez que programava um computador com uma interface gráfica de usuário, e a primeira vez que trabalhava em aplicações grandes e complexas. Escrevia software para um sistema operacional criado pela Grid Systems. Com janelas, fontes múltiplas e menus, o sistema operacional da Grid era muito avançado para sua época.

Em um dia ocorreu-me que o que eu fazia raiava no impossível. Como programador escrevia uma por uma as linhas de código. Agrupava as linhas de código em blocos chamados sub-rotinas. As sub-rotinas eram agrupadas em módulos. Os módulos eram combinados para formar uma aplicação. Os programas de planilha nos quais eu estava trabalhando tinham tantas sub-rotinas e módulos que nenhuma pessoa seria capaz de entender tudo. Era complexo. Não obstante, uma única linha de código fazia muito pouco. Colocar um pixel no monitor de vídeo levava várias linhas de código. Desenhar uma tela inteira para a planilha requeria que o computador executasse milhões de instruções que se estendiam em centenas de sub-rotinas. As sub-rotinas exigiam outras sub-rotinas de forma repetitiva. Era tão complexo que acabava sendo impossível saber tudo o que aconteceria quando o programa estivesse em funcionamento. Ocorreu-me que era muito pouco provável que quando o programa funcionasse eu obtivesse sua imagem na qual apareceria em um instante. Sua aparência externa eram tabelas de números, rótulos, texto e gráficos. Comportava-se como uma planilha. Mas eu sabia o que acontecia dentro do computador, cujo processador executava uma a uma instruções simples. Era difícil pensar que ele pudesse encontrar o caminho pelo labirinto de módulos e sub-rotinas, e executar todas essas instruções tão rápido. Se não o tivesse conhecido tão bem, teria estado seguro de que o conjunto não poderia funcionar. Dava-me conta de que se alguém tivesse inventado o conceito de computador com interface gráfica de usuário e uma aplicação de planilha, e me tivesse apresentado em papel, o teria recusado como algo irreal. Teria afirmado que demoraria uma eternidade para fazer qualquer coisa. Era um pensamento ofensivo, porque isto, de fato, funcionava. Foi então que me dei conta de que meu sentido intuitivo sobre a velocidade do microprocessador e a potência do projeto hierárquico era inadequado.

Há nisso uma lição sobre o neocórtex. Ele não está formado por componentes super-rápidos e as regras com as quais opera não são tão complexas. No entanto, ele apresenta uma estrutura hierárquica que contém bilhões de neurônios e trilhões de sinapses. Se custa-nos imaginar como esse sistema de memória, simples desde a perspectiva lógica mas vasto desde a numérica, pode criar nossa consciência, linguagens, culturas, arte, este livro e nossa ciência e tecnologia, sugiro que isso se deve por que nosso sentido intuitivo sobre a capacidade do

córtex e a potência da sua estrutura hierárquica é inadequado. E, como um computador, acabaremos construindo máquinas inteligentes que funcionem nos mesmos princípios.

Consciência e Criatividade

Quando dou palestras sobre a minha teoria do cérebro, o público costuma captar em seguida o significado da predição como algo unido com a multidão de atividades humanas, e me formulam perguntas relacionadas. De onde provém a criatividade? O que é a consciência? O que é a imaginação? Como se pode separar a realidade das falsas crenças? Ainda que estes temas não tenham estado entre minhas motivações prioritárias para estudar os cérebros, são de interesse para quase todo mundo. Não pretendo ser um expert neles, mas o modelo de memória-predição da inteligência pode contribuir em algumas respostas e percepções úteis. Neste capítulo abordo algumas das perguntas mais frequentes.

São Inteligentes os Animais?

É um rato inteligente? É um gato inteligente? Quando começou a inteligência no período evolutivo? Gosto destas perguntas porque a resposta acaba me sendo surpreendente.

Tudo o que tenho escrito até o momento sobre o neocórtex e seu funcionamento se baseia em uma premissa muito básica: o mundo possui uma estrutura e, portanto, é predizível. Há padrões no mundo: os rostos têm olhos, os olhos têm pupilas, o fogo é quente, a gravidade faz com que os objetos caiam, as portas se abrem e se fecham, e assim sucessivamente. O mundo não é aleatório, nem também não-homogêneo. A memória, a predição e o comportamento não fariam sentido se o mundo carecesse de estrutura. Todo comportamento, seja o de um ser humano, um caracol, um organismo unicelular ou uma árvore, é um meio de explorar a estrutura do mundo em benefício da reprodução.

Imaginemos um organismo unicelular que vive em um açude. A célula tem um flagelo que lhe permite nadar. Na superfície da célula há moléculas que detectam a presença de nutrientes. Como nem todas as zonas do açude apresentam a mesma concentração de nutrientes, se dá uma mudança gradual no valor —ou gradiente— dos nutrientes de um lado da célula ao outro. Quando ela nada pelo açude, a célula pode detectar a mudança. É uma forma de estrutura simples no mundo do animal unicelular. A célula explora sua percepção química nadando para os lugares com concentrações mais altas e nutrientes. Caberia dizer que este organismo simples está efetuando uma predição. Prediz ele que nadando do modo certo achará mais nutrientes. Está a memória envolvida nesta predição? Sim, está. A memória está no DNA do organismo. O animal unicelular não aprendeu em sua vida a explorar este gradiente, particularmente, a aprendizagem ocorreu no período evolutivo e está armazenado em seu DNA. Se a estrutura do mundo mudasse de repente, este animal unicelular particular não aprenderia a se adaptar. Não poderia alterar seu DNA ou o comportamento resultante. Para esta espécie a

aprendizagem só ocorre mediante processos evolutivos ao longo de muitas gerações.

É inteligente este organismo unicelular? Utilizando a noção cotidiana de inteligência humana, a resposta é não. Mas o animal encontra-se no limite extremo de um contínuo de espécies que usam a memória e a predição para conseguir se reproduzir melhor, e segundo esta avaliação mais acadêmica a resposta é sim. Não se trata de etiquetar a algumas espécies como inteligentes e a outras como não inteligentes. Todos os seres vivos usam a memória e a predição. Não há mais do que um contínuo de métodos e complexidade em seu modo do fazer.

As plantas também empregam a memória e a predição para explorar a estrutura do mundo. Uma árvore efetua uma predição quando envia suas raízes sob o solo e seus ramos e folhas para o céu. A árvore prediz onde vai encontrar água e minerais baseando na experiência dos seus antepassados. Certamente, uma árvore não pensa; seu comportamento é automático. Mas a espécie explora a estrutura do mundo do mesmo modo que o organismo unicelular. Cada espécie de plantas possui um conjunto característico de comportamentos que exploram partes ligeiramente diferentes da estrutura do mundo.

As plantas acabaram desenvolvendo sistemas de comunicação baseados em boa medida na libertação lenta de sinais químicos. Se um inseto danifica uma parte de uma árvore, esta envia produtos químicos através do seu sistema vascular a outras partes, o que desencadeia um sistema de defesa como a fabricação de toxinas. Mediante esse sistema de comunicação, a árvore pode mostrar um comportamento um pouco mais complexo. É provável que os neurônios evoluíssem como um meio de comunicar informação mais rápido do que um sistema vascular vegetal. Caberia conceber um neurônio como uma célula com seus próprios apêndices vasculares. Em um determinado momento, em vez de transferir lentamente produtos químicos através desses apêndices, o neurônio começou a utilizar picos eletroquímicos que viajavam com muito mais velocidade. É provável que a princípio a transmissão sináptica rápida e os sistemas nervosos simples não necessitassem de muita aprendizagem. O jogo consistia só em transmitir com maior rapidez.

Mas depois, em decorrência do período evolutivo, aconteceu algo muito interessante. As conexões entre os neurônios tornaram-se modificáveis. Um neurônio podia enviar ou não um sinal dependendo do que tivesse acontecido há pouco tempo. Agora o comportamento podia modificar-se no curso da vida de um organismo. O sistema nervoso tornou-se plástico, assim como o comportamento. Como se podiam formar memórias com rapidez, o animal era capaz de aprender a estrutura do seu mundo durante sua própria vida. Se o mundo mudasse de improviso —digamos que um novo predador entrasse em cena—, o animal não teria que prosseguir com seu comportamento determinado geneticamente, pois talvez já não fosse o apropriado. Os sistemas nervosos plásticos converteram-se

em uma tremenda vantagem evolutiva que levou ao surgimento de novas espécies, dos peixes aos caracóis e dos caracóis aos humanos.

Como temos visto no capítulo 3, todos os mamíferos possuem um cérebro velho, em cima do qual se assenta o neocórtex, que não é mais do que o tecido neural de evolução mais recente. Mas com sua estrutura hierárquica, representações invariáveis e predições por analogia, o córtex permite aos mamíferos explorar bem mais a estrutura do mundo do que é capaz um animal que não o possua. Nossos antepassados dotados de córtex podiam imaginar como fazer uma rede para apanhar peixes. Os peixes não são capazes de aprender que as redes significam morte nem de pensar em como construir ferramentas para as cortar. Todos os mamíferos, dos ratos aos gatos e dos gatos aos humanos, têm neocórtex. Todos são inteligentes, mas em graus diferentes.

O Que é Diferente na Inteligência Humana?

O modelo de memória-predição oferece duas respostas a esta pergunta. A primeira é bastante direta: nosso neocórtex é maior que o do macaco ou do cão, por exemplo. Ao ampliar a lâmina cortical até o tamanho de um grande guarda-roupa, nossos cérebros podem aprender um modelo mais complexo do mundo e efetuar predições mais complexas. Vemos analogias mais profundas, mais estrutura na estrutura, do que outros mamíferos. Se desejamos encontrar um colega, não nos limitamos a olhar atributos simples como a saúde, mas entrevistamos a seus amigos e pais, observamos como eles se comportam e falam, e julgamos quão honestos eles são. Observamos estes atributos secundários e terciários para tentar prever qual será o comportamento do nosso colega no futuro. Os corredores das bolsas de valores buscam uma estrutura nos padrões do mercado. Os matemáticos buscam uma estrutura nos números e equações. Os astrônomos buscam uma estrutura nos movimentos dos planetas e das estrelas. Nosso neocórtex permite-nos contemplar nossa casa como parte de uma cidade, que é parte de uma região, que é parte de um planeta, o que é parte de um grande Universo: a estrutura dentro da estrutura. Nenhum outro mamífero pode ponderar a esta profundidade. Estou seguro de que minha gata não tem um conceito do mundo fora da nossa casa.

A segunda diferença entre os seres humanos e o restante dos mamíferos é que temos linguagem. Escreveram-se livros inteiros sobre as supostas propriedades únicas da linguagem e como ela se desenvolveu. No entanto, ela encaixa-se perfeitamente no modelo de memória-predição sem nenhum molho especial nem maquinaria específica. As palavras faladas e escritas não são mais do que padrões do mundo, assim como as melodias, os carros e as casas. A sintaxe e a semântica da linguagem não são diferentes da estrutura hierárquica de outros objetos cotidianos. E do mesmo modo que associamos o som de um trem com a imagem da memória visual de um trem, associamos as palavras faladas com nossa memória dos seus homólogos físicos e semânticos. Mediante a linguagem um ser

humano pode evocar lembranças e criar novas justaposições de objetos mentais em outro ser humano. A linguagem é analogia pura, e através dela podemos conseguir que outros seres humanos experimentem e aprendam coisas que talvez não tenham visto nunca. O desenvolvimento da linguagem requereu um grande neocórtex capaz de manejar a estrutura aninhada da sintaxe e da semântica. Requereu, além disso, um córtex motor e uma musculatura mais plenamente desenvolvidos que nos permitissem realizar sons ou gestos sofisticados e muito articulados. Com a linguagem podemos tomar os padrões que aprendemos na vida e transmiti-los a nossos filhos e nossa clã. A linguagem, seja escrita, falada ou incorporada nas tradições culturais, converteu-se no meio com o qual transmitimos o que sabemos do mundo de geração em geração. Hoje a comunicação impressa e eletrônica permite-nos compartilhar nosso conhecimento com milhões de pessoas ao redor do mundo. Os animais carentes de linguagem não transmitem tanta informação a sua progênie. Um rato pode aprender muitos padrões em sua vida, mas não transmite nova informação em detalhes: "Olha, filho, assim é como me ensinou meu pai a evitar as descargas elétricas".

Portanto, podem-se determinar três etapas na inteligência, empregando-se em todas elas a memória e a predição. A primeira seria quando as espécies se valiam do DNA como meio para estabelecer a memória. Os indivíduos não podiam aprender e se adaptar dentro da sua vida. Só eram capazes de transmitir à sua descendência a memória do mundo baseada no DNA através dos seus genes.

A segunda etapa iniciou-se quando a Natureza inventou sistemas nervosos modificáveis, capazes de formar memórias com rapidez. Agora um indivíduo conseguia aprender coisas importantes sobre a estrutura do seu mundo e adaptar seu comportamento durante sua vida. Mas ainda não tinha a faculdade de comunicar este conhecimento a sua prole mais do que pela observação direta. A criação e expansão do neocórtex ocorreu durante esta segunda etapa, mas não a definiu.

A terça e última etapa é única dos seres humanos. Começa com a invenção da linguagem e a expansão do nosso grande neocórtex. Nós os humanos podemos aprender muito da estrutura do mundo dentro das nossas vidas e o comunicar com efetividade a muitos outros humanos através da linguagem. Você e eu estamos participando neste processo agora mesmo. Tenho passado grande parte da minha vida indagando na estrutura dos cérebros e em como a dita estrutura leva ao pensamento e à inteligência. Mediante este livro estou difundindo o que tenho aprendido. Certamente, não poderia o ter feito se não tivesse tido acesso ao conhecimento reunido por milhares de cientistas, que aprenderam de outros, e assim sucessivamente através dos séculos. Fui capaz de assimilar e ampliar o que outros tinham escrito sobre sua reflexão e observação.

Convertemo-nos nas criaturas mais adaptáveis do planeta e as únicas com a capacidade de transferir amplamente nosso conhecimento do mundo a nossa

espécie. A população humana tem passado por um crescimento explosivo porque temos a faculdade de aprender e explorar a estrutura do mundo e comunicá-la a outros seres humanos. Podemos prosperar em qualquer lugar, seja a selva tropical chuvosa, o deserto ou a tundra gelada. A combinação de um neocórtex grande e a linguagem tem conduzido à espiral de sucesso da nossa espécie.

O Que é a Criatividade?

Perguntam-me com frequência sobre a criatividade, suspeito que é porque muitas pessoas a consideram algo que uma máquina não poderia fazer e, portanto, constitui um desafio para a ideia de construir máquinas inteligentes. O que é a criatividade? Já temos encontrado a resposta várias vezes neste livro. A criatividade não é algo que ocorre em uma região particular do córtex; também não se parece com as emoções ou com o equilíbrio, que têm sua origem em estruturas e circuitos particulares fora do córtex. A criatividade é mais do que uma propriedade inerente de cada região cortical. É um componente necessário da predição.

Como pode ser verdade isto? Não se trata de uma qualidade necessária que requer inteligência e talento elevados? A verdade é que não. A criatividade pode-se definir claramente como a faculdade de elaborar predições por analogia, algo que ocorre em todas partes do córtex e que fazemos de forma contínua enquanto estamos acordados. A criatividade ocorre em um contínuo. Abrange desde os atos quotidianos simples de percepção que acontecem nas regiões sensoriais do córtex (escutar uma canção em um novo tom), até os atos difíceis e raros de gênio que têm lugar nos níveis superiores do córtex (compor uma sinfonia em um estilo novo). No fundamental, os atos quotidianos de percepção são similares aos raros voos de brilhantismo; o que acontece é que os atos quotidianos são tão comuns que não lhes damos importância.

Agora você já conta com um entendimento básico de como criamos memórias invariáveis, como as empregamos para efetuar predições e como fazemos predições de acontecimentos futuros que sempre são um pouco diferentes do que temos experimentado no passado. Recorde também que nossas memórias invariáveis são sequências de fatos. Efetuamos predições combinando a lembrança da memória invariável do que deverá acontecer a seguir com os detalhes pertencentes do momento temporário (recordemos a parábola da predição sobre a chegada do trem). A predição é a aplicação de sequências de memória invariável à novas situações. Portanto, todas as predições corticais o são por analogia. Predizemos o futuro por analogia ao passado.

Consideremos que estamos a ponto de comer em um restaurante desconhecido e queremos lavar as mãos. Ainda que jamais tenhamos estado nesse edifício antes, nosso cérebro prediz que terá banheiro em algum lugar do estabelecimento com uma pia apropriada para se lavar as mãos. Como se sabe isto? Outros restaurantes

nos quais tínhamos estado têm serviços e por analogia é provável que este também os tenha. E além disso, sabemos onde e o que buscar. Predizemos que terá uma porta ou cartaz com algum tipo de símbolo associado à homens ou mulheres. Predizemos que estará na parte traseira do restaurante, próximo ao bar ou à entrada, mas em geral fora da vista da zona de mesas. Novamente, nunca tínhamos estado neste restaurante particular antes, mas por analogia com outros estabelecimentos similares somos capazes de encontrar o que precisamos. Não olhamos ao redor de forma aleatória. Buscamos padrões esperados que nos permitam encontrar os serviços em seguida. Este tipo de comportamento é um ato criativo; é predizer o futuro por analogia com o passado. Não costumamos pensar que isto seja criativo, mas é, e muito.

Há um tempo atrás eu havia comprado um vibrafone. Tínhamos piano, mas nunca tinha tocado vibrafone antes. No dia que o levamos a casa, apanhei uma partitura do piano, a coloquei no atril sobre o vibrafone e comecei a tocar melodias simples. Minha habilidade para fazer isto não era notável, mas no básico era um ato criativo. Pensemos sobre minha atividade. Tenho um instrumento que é muito diferente do piano. O vibrafone possui barras de metal douradas; o piano, teclas negras e brancas. As barras douradas são grandes e todas do mesmo tamanho; as teclas, pequenas e de dois tamanhos diferentes. As barras douradas estão dispostas em duas filas diferentes; as teclas negras e brancas, intercaladas. Em um instrumento emprego os dedos; no outro, as baquetas. Em um caso, permaneço de pé; no outro, sentado. Os músculos e movimentos particulares necessários para tocar o vibrafone são totalmente diferentes dos que são precisos para tocar o piano.

Portanto, como eu soube tocar uma melodia em um instrumento desconhecido? A resposta é que meu córtex vê uma analogia entre as teclas do piano e as barras do vibrafone. O emprego desta semelhança permitiu-me interpretar a melodia. A verdade é que não é diferente de cantar uma canção em um novo tom. Em ambos casos sabemos o que fazer por analogia com a aprendizagem passada. Dou-me conta de que a semelhança entre estes dois instrumentos pode lhe parecer evidente, mas isso se deve por que nossos cérebros veem analogias de forma automática. Tente programar um computador para que encontre semelhanças entre dois objetos como pianos e vibrafones, e comprovarão o quão difícil é. A predição por analogia —a criatividade— é tão dominante que não costumamos a perceber.

No entanto, achamos que somos criativos quando nosso sistema de memória-predição opera a um nível de abstração superior, quando efetua predições pouco comuns utilizando analogias infrequentes. Por exemplo, a maioria das pessoas estaria de acordo de que o matemático que demonstra uma conjectura difícil é criativo. Mas observemos detidamente o que implicam seus esforços mentais. Nosso matemático concentra-se muito em uma equação e afirma: “Como vou

desvendar este problema?”. Se a resposta não é evidente à primeira vista, ele talvez reordene a equação. Ao escrevê-la de forma diferente, ele pode contemplar o mesmo problema desde uma perspectiva diferente. Ele concentra-se um pouco mais. De improviso ele vê uma parte da equação que lhe parece conhecida. Ele pensa: “Epa, reconheço isto. Há uma estrutura nesta equação que se parece à estrutura de outra equação em que trabalhei há vários anos”. Então ele efetua uma predição por analogia: “Talvez possa resolver esta equação utilizando as mesmas técnicas que me serviram no caso da antiga”. Ele é capaz de resolver o problema por analogia com um problema aprendido previamente. É um ato criativo.

Meu pai padecia de um misterioso transtorno no sangue que seus médicos não conseguiam diagnosticar. Como souberam que tratamento oferecer? Uma das coisas que fizeram foi analisar meses de dados obtidos da análise do seu sangue para poderem identificar padrões. (Meu pai imprimiu um belo diagrama para que os médicos pudessem ver os dados com clareza.) Ainda que seus sintomas não se encaixassem plenamente com os das doenças conhecidas, existia algumas semelhanças. Os médicos acabaram baseando seu tratamento em uma mistura de estratégias das quais tinham obtido bons resultados em outros transtornos sanguíneos. Os tratamentos empregados deduziram-se a partir de analogias com doenças que os médicos tinham tratado antes. Para reconhecer estes padrões requer-se um amplo contato com outras doenças pouco comuns.

As metáforas de Shakespeare são um exemplo da criatividade. “O amor é uma fumaça criada com o vapor dos suspiros”; “filosofia, doce leite da adversidade”; “há adagas nos sorrisos dos homens”. Tais metáforas parecem óbvias quando as vemos, mas são muito difíceis de inventar, razão pela qual se considera a Shakespeare um gênio literário. Para criar estas metáforas ele teve que ver uma sucessão de analogias inteligentes. Quando ele escreve “há adagas nos sorrisos dos homens”, ele não fala de adagas ou sorrisos. Adagas é análoga a más intenções, e os sorrisos dos homens, a engano. Duas sábias analogias em apenas cinco palavras! Pelo menos é minha forma de interpretá-las. Os poetas têm o dom de correlacionar palavras ou conceitos que não parecem ligados, de forma que dão uma nova luz ao mundo. Criam analogias inesperadas como meio para ensinar uma estrutura de nível superior.

De fato, as obras de arte muito criativas são apreciadas porque violam nossas predições. Quando vemos um filme que rompe o molde conhecido de um personagem, argumento ou cinematografia (incluindo efeitos especiais), gostamos porque não é o mesmo de sempre. A pintura, a música, a poesia e as novelas — todas as formas artísticas criativas — se esforçam para romper com o convencional e violar as expectativas do público. Existe uma tensão contraditória que engrandece uma obra de arte. Queremos que a arte tenha um ar familiar, mas ao mesmo tempo que seja único e inesperado. Se é muito conhecido, acaba sendo recauchutado ou *kitsch*; muita singularidade faz com que desentone e seja difícil

de apreciar. As melhores obras rompem alguns padrões esperados e ao mesmo tempo nos ensinam outros novos. Observe uma grande obra de música clássica. A melhor música acaba sendo atraente por sua simplicidade —bom compasso, melodia e fraseologia simples—. Qualquer pessoa pode a entender e a apreciar. No entanto, também é um pouco diferente e inesperada. Mas quanto mais escutamos ela, mais vemos que há um padrão nas partes inesperadas, como harmonias pouco habituais ou repetidas mudanças de tons. O mesmo cabe afirmar da grande literatura ou dos grandes filmes. Quanto mais os lemos ou os assistimos, maior detalhe criativo e complexidade de estrutura percebemos.

É provável que tenham tido a experiência de estar olhando algo e de repente lhe vir à cabeça uma sensação de reconhecimento: “Tenho visto este padrão antes, em algum outro lugar...”. Talvez não estivesse tentando resolver um problema, mas uma representação invariável do seu cérebro se ativou por uma situação nova. Você vê uma analogia entre dois acontecimentos que não costumam estar relacionados. Talvez eu devesse reconhecer que promover uma ideia científica é similar a vender uma ideia comercial ou que propiciar a reforma política é semelhante a criar filhos. Se eu fosse poeta, teria uma nova metáfora. Se eu fosse cientista ou engenheiro, teria uma nova solução para um problema persistente. A criatividade consiste em misturar e casar padrões de tudo o que temos experimentado ou chegado a conhecer em nossa vida. É afirmar: “Isto se parece com isso”. O mecanismo neural para se fazer isto encontra-se em todas as partes do córtex.

São Algumas Pessoas Mais Criativas Que Outras?

Uma pergunta relacionada que costumo escutar é: “Se todos os cérebros são inerentemente criativos, por que existem diferenças na nossa criatividade?”. O modelo de memória-predição aponta duas respostas possíveis. Uma tem a ver com a Natureza, e a outra, com a educação.

No aspecto educativo, cada pessoa tem diferentes experiências de vida. Portanto, cada pessoa desenvolve diferentes modelos e memórias do mundo em seu córtex, e efetuarão analogias e predições diferentes. Se tenho tido contato com a música, serei capaz de cantar uma canção em um novo tom e tocar melodias simples em novos instrumentos. Se nunca tive contato com ela, não conseguirei realizar esses saltos preditivos. Se tenho estudado física, serei capaz de explicar o comportamento dos objetos quotidianos mediante a analogia com as leis da física. Se tenho crescido com cães, estou preparado para ver analogias com eles e conseguirei prever melhor suas comportamentos. Algumas pessoas são mais criativas em situações sociais ou em linguagem, matemática ou diplomacia segundo o ambiente no qual cresceram. Nossas predições e, portanto, nossos talentos constroem-se partindo das nossas experiências.

No capítulo 6 tenho descrito como as memórias se impulsionam para abaixo na hierarquia cortical. Quanto mais expomo-nos a certos padrões, mais são reformadas as memórias dos ditos padrões em níveis inferiores. Isto nos permite aprender a relação entre os objetos abstratos de ordem superior que se encontram no topo. É a essência da perícia. Um expert é alguém que mediante a prática e a exposição repetida é capaz de reconhecer padrões mais sutis dos que consegue determinar alguém não-expert, como a forma de uma asa em um carro do final da década de 1950 ou o tamanho de uma mancha no bico de uma gaivota. Os experts podem reconhecer padrões sobre padrões. O limite físico do que somos capazes de reconhecer é delimitado pelo tamanho do nosso córtex. Mas, como seres humanos, nosso córtex é grande comparado com outras espécies e gozamos de uma tremenda flexibilidade na qual podemos aprender. Tudo depende das coisas com as quais temos contato ao longo da nossa vida.

No aspecto natural, os cérebros mostram uma variação física. Sem dúvida, algumas das diferenças estão determinadas pelos genes, como o tamanho das regiões (indivíduos podem mostrar uma diferença de até três vezes na área bruta V1) e a lateralidade hemisférica (as mulheres tendem a ter um cabeamento mais denso ligando os lados esquerdo e direito do cérebro que os homens). Entre os humanos, é provável que alguns cérebros apresentem mais células ou diferentes tipos de conexões. Não parece que o gênio criativo de Albert Einstein se desse apenas em função do ambiente estimulante do escritório de patentes em que ele trabalhou quando jovem. As análises recentes do seu cérebro —que se acreditava estar perdido, mas foi encontrado conservado em um jarro há alguns anos— revelam que ele era sensivelmente inusual. Tinha mais células de apoio — chamadas neuróglias— por neurônio que a média. Mostrava um padrão pouco habitual de cisuras ou sulcos nos lóbulos parietais, região que se considera importante para as capacidades matemáticas e o raciocínio espacial. Além disso, era 15 por cento mais amplo que a maioria dos cérebros restantes. Talvez nunca saibamos por que Einstein foi tão criativo e inteligente, mas cabe apostar com total segurança que parte do seu talento se derivou de fatores genéticos.

Seja qual for a diferença entre cérebros brilhantes e normais, todos somos criativos. E com a prática e o estudo podemos aumentar nossas destrezas e talentos.

Podemos Treinar-nos para Ser Mais Criativos?

Sim, sem sombra de dúvidas. Tenho descoberto que há três modos de fomentar o encontro de analogias úteis quando se trabalha na resolução de problemas. Primeiro é preciso presumir que há uma resposta ao que tratamos de resolver. As pessoas rendem-se com muita facilidade. Precisamos confiar em que há uma solução esperando a ser descoberta e devemos continuar pensando no problema durante um bom tempo.

Em segundo lugar, temos que deixar vagar nossa mente. Temos de dar a nosso cérebro o tempo e o espaço precisos para descobrir a solução. Encontrar a solução a um problema é, de forma literal, encontrar um padrão armazenado em nosso córtex que é análogo ao problema no qual estamos trabalhando. Se perseveramos no problema, o modelo de memória-predição sugere que devemos encontrar modos diferentes do considerar para aumentar a probabilidade de ver uma analogia com uma experiência passada. Se limitamo-nos a sentar e olhá-lo fixamente uma e outra vez, não chegaremos muito longe. Temos de tentar tomar as partes do problema e reordená-las de forma literal e figurada. Quando jogo Scrabble, mudo uma e outra vez a ordem das fichas. Não que eu espere que as letras formem por acaso uma nova palavra, mas combinações diferentes farão com que eu me lembre de palavras ou pedaços de palavras que poderiam ser parte de uma solução. Se você olha um desenho de algo que não faz sentido, tente girá-lo, mudar as cores ou as perspectivas. Por exemplo, quando eu pensava sobre como os diferentes padrões de V1 poderiam levar a representações variáveis em IT, eu fiquei travado, de modo que dei a volta ao problema para me perguntar como um padrão constante de IT era capaz de conduzir a predições diferentes em V1. O invertimento do problema acabou sendo útil de imediato e acabou levando-me a achar que V1 não devia se considerar uma região cortical única.

Se você fica travado em um problema, abandone por um tempo. Faça outra coisa. Depois retome-no, propondo-o novamente. Se você fizer isto o suficiente, algo surgirá mais cedo ou mais tarde. Pode levar dias ou semanas, mas acabará acontecendo. A meta é encontrar uma situação análoga em algum lugar da sua experiência passada. Para conseguir isto você deve ponderar sobre o problema com frequência, mas também fazer outras coisas a fim de que o córtex tenha a oportunidade de encontrar uma memória análoga.

Vejam os outros exemplos de como a reordenação de um problema levou a uma solução inovadora. Em 1994 meus colegas e eu tentávamos criar uma forma de introduzir texto em computadores de mão. Todos nós estávamos centrados no *software* de reconhecimento de letra. Dizíamos: "Escrevemos coisas em folhas de papel, de modo que devemos ser capazes de fazer o mesmo na tela de um computador". Infelizmente, isto acabou sendo difícilíssimo. É outra dessas coisas que os computadores não fazem nada bem, ainda que aos cérebros lhes pareça muito simples. A razão é que o cérebro usa a memória e o texto em curso para prever o que está escrito. Palavras e letras irreconhecíveis em si mesmas se reconhecem com facilidade no contexto. No caso dos computadores não basta casar padrões para realizar a tarefa. Eu tinha projetado vários computadores que empregavam reconhecimento de letra tradicional, mas nunca tinham sido muito bons.

Durante vários anos empenhei-me em fazer com que o *software* de reconhecimento funcionasse melhor, porém eu ficava travado. Em um dia decidi

afastar-me um pouco e contemplar o problema de uma perspectiva diferente. Busquei problemas análogos. Disse a mim mesmo: "Como introduzimos texto nos computadores de mesa? O digitamos em um teclado. Como sabemos como é digitar em um teclado? Bom, a verdade é que não é fácil. É uma invenção muito recente e leva muito tempo para aprender. Escrever com o toque em um teclado parecido com o de uma máquina de escrever é difícil e não é intuitivo; não se parece de forma alguma a escrever, mas milhões de pessoas aprendem a fazer. Por que? Porque funciona". Meu pensamento continuou por analogia: "Talvez eu possa criar um sistema de entrada de texto que não seja necessariamente intuitivo, e que se tenha que aprender, mas que as pessoas o utilizem porque funciona".

Esse é literalmente o processo que segui. Utilizei o ato de escrever em um teclado como analogia para criar uma forma de inserir texto com um ponteiro em uma tela. Reconheci que as pessoas estavam dispostas a aprender uma tarefa difícil (digitar) porque era um modo confiável e fácil de introduzir texto em uma máquina. Portanto, se éramos capazes de criar um novo método de introduzir texto com um ponteiro que fosse rápido e confiável, as pessoas o utilizariam ainda que requeresse aprendizagem. Portanto, projetei um alfabeto que traduzisse de forma confiável o que escrevíamos em texto computacional, e o chamei de Graffiti. Nos sistemas de reconhecimento de letra tradicionais, quando o computador interpreta mal nossa escrita não sabemos por que. Mas o sistema Graffiti sempre produz a letra correta a não ser que cometamos um erro ao escrever. Nossos cérebros detestam o que não é predizível, motivo pelo qual as pessoas sentem aversão pelos sistemas de reconhecimento de letra tradicionais.

Muita gente pensou que Graffiti era uma ideia completamente estúpida. Ia na contramão de tudo o que se acreditava sobre como se supunha que funcionavam os computadores. Naqueles dias o mantra era que os computadores tinham que se adaptar ao usuário e não o contrário. Mas eu confiava em que as pessoas aceitariam este novo modo de introduzir texto por analogia com o teclado. Graffiti acabou sendo uma boa solução e sua adoção foi ampla. Até hoje sigo ouvindo gente declarar que os computadores devem adaptar aos usuários, mas nem sempre isto é verdade. Nossos cérebros preferem sistemas coerentes e predizíveis, e gostamos de aprender novas destrezas.

Pode a Criatividade Me Desencaminhar? Posso Enganar a Mim Mesmo?

A falsa analogia é sempre um perigo. A história da ciência está repleta de belas analogias que acabaram sendo errôneas. Por exemplo, o famoso astrônomo Johannes Kepler ficou convencido de que as órbitas dos seis planetas conhecidos estavam definidas pelos sólidos platônicos. Estas são as únicas formas tridimensionais que podem ser completamente construídas partindo de polígonos regulares. Existem só cinco: tetraedro (quatro triângulos equiláteros), hexaedro (seis quadrados, também conhecido como cubo), octaedro (oito triângulos equiláteros), dodecaedro (doze pentágonos regulares) e icosaedro (vinte triângulos

equiláteros). Elas foram descobertas pelos gregos antigos, que estavam obsecados com a relação da matemática com o Cosmos.

Assim como todos os eruditos renascentistas, Kepler estava muito influenciado pelo pensamento grego. Parecia-lhe que não podia ser uma coincidência que tivesse cinco sólidos platônicos e seis planetas. Assim ele expressa isto em seu livro *O mistério do Cosmos* (1596): "O mundo dinâmico está representado pelos sólidos de faces planas. Há cinco, mas quando os considera como limites, esses cinco determinam seis coisas diferentes: daí os seis planetas que giram ao redor do Sol. Também é esta a razão pela qual não há mais do que seis planetas". Ele viu uma analogia formosa, mas completamente falsa.

Kepler prosseguia explicando as órbitas dos planetas recorrendo à hierarquização dos sólidos platônicos com o Sol no centro. Ele tomou a esfera definida pela órbita de Mercúrio como linha de referência e circunscreveu nela um octaedro cujas pontas definiam uma esfera maior, que proporcionava a órbita de Vênus. Ao redor desta ele circunscreveu um icosaedro, cujas pontas exteriores produziam a órbita da Terra. A progressão continuava: o dodecaedro projetado ao redor da órbita terrestre dava a órbita de Marte; o tetraedro em torno desta apontava a órbita de Júpiter, e o cubo ao redor desta indicava a órbita de Saturno. Era elegante e belo. Dada a precisão limitada dos dados astronômicos da sua época, ele pôde convencer-se de que este esquema funcionava. (Anos mais tarde, Kepler deu-se conta de que tinha se equivocado depois dele se apoderar dos dados astronômicos precisos do seu colega falecido Tycho Brahe, que demonstravam que as órbitas planetárias eram elipses e não círculos.)

A exaltação de Kepler serve como lição de moral para os cientistas e, na realidade, para todos os pensadores. O cérebro é um órgão que constrói modelos e efetua previsões criativas, mas seus modelos e previsões podem ser tanto enganosos como válidos. Nossos cérebros sempre estão buscando padrões e fazendo analogias. Se não se podem encontrar correlações corretas, o cérebro se mostra mais do que contente em aceitar as falsas. A pseudociência, o fanatismo, a fé e a intolerância costumam ter suas raízes na falsa analogia.

O que é a Consciência?

É uma dessas perguntas que temem os neurocientistas, em minha opinião de forma desnecessária. Alguns cientistas, como Christof Koch, estão dispostos a desvendar o tema da consciência, mas a maioria a considera um assunto da filosofia que bordeia a pseudociência. Penso que merece consideração ainda que seja apenas isto, porque desperta a curiosidade de muita gente. Não posso proporcionar uma resposta completamente satisfatória, mas acho que a memória e a previsão a abordam em parte. Em primeiro lugar, falemos dessa charada tal como surge na conversa.

Não faz muito tempo que me encontrava em uma conferência científica em um belo lugar de Long Island Sound. Nas primeiras horas da tarde, alguns de nós apanhamos nossos copos de vinho e os levamos até um cais para sentar-nos junto à água e para bater papo antes da comida e da sessão vespertina. Decorrido um momento, a conversa girou em torno do assunto da consciência. Como já disse, os neurocientistas não costumam falar disso, mas estávamos em um bonito lugar, tínhamos bebido vinho e surgiu o tema.

Uma cientista britânica soltou uma conclusão sobre suas ideias a respeito e afirmou: “Certamente, jamais compreenderemos a consciência”. Eu não estava de acordo: “A consciência não é um grande problema. Acho que não é mais do que o que se sente ao ter córtex”. Fez-se um silêncio no grupo e depois suscitou-se uma discussão quando vários cientistas tentaram ilustrar sobre meu erro evidente. “Você deve admitir que o mundo parece vivo e formoso. Como pode negar que sua consciência percebe o mundo? Você deve admitir que se sente como algo especial.” Para estabelecer minha premissa, repliquei: “Não sei do que falam. Dado o modo em que estão se expressando sobre a consciência, tenho que concluir que sou diferente de vocês. Não sinto o que vocês estão sentindo, de modo que talvez não seja um ser consciente. Devo ser um zumbi”. Costuma-se recorrer aos zumbis quando os filósofos falam sobre a consciência. Um zumbi define-se como alguém fisicamente igual à um humano, porém sem consciência. São máquinas de carne e osso que caminham e respiram, mas sem ninguém dentro.

A cientista britânica olhou-me. “Claro que é você consciente.”

“Não, não o creio. É provável que pareça, mas não sou um ser humano consciente. Não se preocupe por isso; encontro-me a gosto.”

Ela disse, “Bom, é que você não percebe a maravilha?” e estendeu o braço para a água cintilante enquanto o sol começava a se por e o céu tornava-se rosa salmão iridescente.

“Sim, vejo todas essas coisas; e daí?”

“Então, como explica sua experiência subjetiva?”

Repliquei, “Sim, sei que estou aqui. Tenho memória de coisas semelhantes a este entardecer. Mas não me parece que esteja acontecendo nada especial, de modo que se você sente algo especial talvez seja porque não sou consciente.” Eu tentava fazer-lhe definir o que ela pensava que tinha de maravilhoso e inexplicável na consciência. Eu tentava fazer com que ela a definisse.

Continuamos nesta linha de raciocínio —sim, o é; não, não sou isso— até que chegou o momento de nos dirigir a comer. Não acho que eu pudesse mudar o modo de pensar de ninguém sobre a existência e o significado da consciência. Mas

tratava de conseguir que se dessem conta de que a maioria das pessoas pensa que ela é uma espécie de molho mágico que se acrescenta sobre o cérebro físico. Temos um cérebro, composto de células, e vertemos a consciência, esse molho mágico, sobre ele, e essa é a condição humana. Nesta proposta, a consciência é uma entidade misteriosa separada dos cérebros. Por isso os zumbis têm cérebros, mas não consciência. Todos possuem todos os elementos mecânicos, neurônios e sinapses, mas carecem do molho especial. Podem fazer tudo o que um humano faz. Pelo exterior não se pode distinguir um zumbi de um humano.

A ideia de que a consciência é algo acrescentado provém de crenças anteriores no *élan* vital, uma força especial que se pensava que animava os seres vivos. As pessoas achavam que precisava-se da dita força para explicar a diferença entre as rochas e as plantas ou entre os metais e as donzelas. Poucos continuam acreditando nisto. Na atualidade sabemos bastante sobre as diferenças entre matéria inanimada e animada para entender que não há um molho especial. Agora sabemos muito sobre DNA, dobramento de proteínas, transcrição de genes e metabolismo. Ainda que ainda não conheçamos todos os mecanismos dos sistemas vivos, sabemos o suficiente de biologia para abandonar a magia. De forma similar, as pessoas já não sugerem que se precisa de magia ou de espíritos para fazer com que os músculos se movam. Temos proteínas que se dobram e formam grandes correntes de moléculas. Pode-se ler de tudo a respeito.

Não obstante, muita gente persiste em achar que a consciência é diferente e não pode se explicar em termos biológicos reducionistas. Reitero que não sou um estudioso da consciência. Não tenho lido todas as opiniões dos filósofos, mas tenho algumas ideias sobre o que confunde as pessoas neste debate. Acho que a consciência é o que se sente ao ter neocórtex. Mas podemos melhorar isto. Podemos dividir a consciência em duas categorias importantes. Uma é similar ao conhecimento de si mesmo, a noção cotidiana de estar consciente, e é relativamente fácil de entender. A segunda são os *qualia*, a ideia de que os sentimentos associados com a sensação são em certo modo independentes da entrada sensorial. É a parte mais difícil.

Quando a maioria das pessoas diz a palavra *consciente* ela está se referindo à primeira categoria. "Você está consciente de que passou a meu lado sem me dizer oi?" "Você estava consciente quando você caiu da cama na noite passada?" "Não está consciente quando você dorme." Algumas pessoas afirmam que esta forma de consciência é o mesmo que o conhecimento. Estão muito próximos em significado, mas não acho que seu conhecimento a capte com acerto. Sugiro que este significado de consciência é sinônimo de formar memórias declarativas. São memórias que se podem recordar e falar delas a outras pessoas. Você pode expressá-las verbalmente. Se você me pergunta para onde fui no fim de semana passado, eu posso lhe dizer. É uma memória declarativa. Se você me pergunta como equilibrar uma bicicleta, eu posso lhe dizer que se deve sustentar a barra do

guidão e empurrar os pedais, mas sou incapaz de lhe explicar com exatidão como o fazer. O modo de equilibrar uma bicicleta tem muito a ver com a atividade do cérebro antigo, de modo que não é uma memória declarativa.

Conto com um pequeno experimento mental para mostrar que nossa noção cotidiana de consciência se corresponde com formar memórias declarativas. Recordemos que se pensa que todas as memórias residem nas mudanças físicas das sinapses e dos neurônios que se ligam. Portanto, se tivesse um método para reverter essas mudanças físicas, nossas memórias se apagariam. Agora imaginemos que você pudesse instalar um interruptor e devolver seu cérebro ao estado físico exato em que ele se achava em algum momento do passado. Poderia ser há uma hora, vinte e quatro horas ou qualquer outra coisa. Só pulso o interruptor em minha máquina de volta ao passado e suas sinapses e neurônios regressam a um estado anterior no tempo. Ao fazer isto, apago toda sua memória do que tem ocorrido desde então.

Suponhamos que você viva o dia de hoje e acorde amanhã. Mas no momento em que está acordando, pulso o interruptor e apago as últimas vinte e quatro horas. Você teria uma memória zero do dia anterior. Desde a perspectiva do seu cérebro, ontem nunca aconteceu. Lhe diria que é quarta-feira e você protestaria: "Não, é terça-feira. Tenho certeza. Você mudou o calendário. É terça-feira, sem dúvida. Por que está fazendo esta brincadeira comigo?". Mas todas as pessoas com as quais você se encontrou na terça-feira diriam que você tinha estado consciente durante todo o dia. Viram-no, comeram e falaram com você. Não se lembra? Você responderia que não; que isto não aconteceu. Por último, quando se lhe mostrasse um vídeo no qual você se vê comendo, você iria se convencer aos poucos de que existiu esse dia, ainda que você não tenha memória dele. É como se você tivesse sido zumbi durante um dia, sem consciência. No entanto, você foi consciente o tempo todo. Sua crença de que você era consciente só desapareceu quando sua memória declarativa se apagou.

Este experimento mental capta a equivalência entre memória declarativa e nossa noção cotidiana de ser conscientes. Se em uma partida de tênis no seu término pergunto-lhe se você é consciente, você me responderia sem dúvida que sim. Se depois apago sua memória das duas últimas horas, você declararia que estava inconsciente e não era responsável por suas ações durante esse tempo. Em ambos casos jogaram a mesma partida de tênis. A única diferença é que você tem memória disso no momento em que pergunto. Portanto, este significado de consciência não é absoluto. Ele pode ser alterado após o fato de ter sido apagada a sua memória.

A questão mais difícil sobre a consciência refere-se aos *qualia*, que se costumam formular em perguntas do tipo zen, como: "Por que o vermelho é vermelho, e o verde, verde? Parece-me o vermelho da mesma forma que parece para você? Por que o vermelho está carregado emocionalmente de certos sentimentos? Sem

dúvida, possui para mim uma qualidade ou sensação inextricável. Que sentimentos ele causa a você?”.

Acho as ditas descrições difíceis de relacionar com a neurobiologia, de modo que gostaria reformular a pergunta. Ao meu entender, uma pergunta equivalente, mas também difícil de explicar, seria por que os diversos sentidos parecem qualitativamente diferentes. Por que a visão parece diferente da audição e por que a audição parece diferente do tato? Se o córtex é o mesmo em todas partes, se funciona com os mesmos processos, se for limitado a manejar padrões, se não entra nenhum som ou luz no cérebro, só padrões, por que a visão parece tão diferente da audição? Acaba-me sendo difícil descrever em que difere a visão da audição, mas é algo evidente. Suponho que também é para vocês. Não obstante, o axônio que representa o som e outro que representa a luz são idênticos em todos os efeitos práticos. As qualidades da luz e do som não se transportam no axônio de um neurônio sensorial.

As pessoas que apresentam uma doença chamada sinestesia tem cérebros que confundem a distinção entre os sentidos: alguns sons ou certas texturas têm cor. Isto nos indica que o aspecto qualitativo de um sentido não é imutável. Mediante algum tipo de modificação física, um cérebro pode conceder um aspecto qualitativo de visão a uma entrada auditiva.

Portanto, qual é a explicação para os *qualia*? Posso pensar em duas possibilidades, mas nenhuma delas me parece totalmente satisfatória. Uma é que ainda que a audição, o tato e a visão funcionem segundo princípios similares no neocórtex, se manejam de forma diferente abaixo deste. A audição descansa em um conjunto de estruturas subcorticais específicas para a audição que processam os padrões auditivos antes que cheguem ao córtex. Os padrões somatosensoriais também viajam por um conjunto de áreas subcorticais que são únicas para os sentidos somáticos. Talvez os *qualia*, como as emoções, não estejam mediados simplesmente pelo neocórtex. Se estão vinculados de algum modo com partes subcorticais do cérebro que possuem uma conexão única, talvez unida com os centros de emoção, isso poderia explicar por que os percebemos de forma diferente, ainda que isto não ajude a resolver por que existe o tipo de sensação de qualia.

A outra possibilidade que me ocorre é que a estrutura das entradas —diferenças entre os mesmos padrões— dite como experimentamos os aspectos qualitativos da informação. A natureza do padrão espacial-temporal no nervo auditivo é diferente da natureza do mesmo padrão no nervo óptico. Este último possui milhões de fibras e transporta bastante informação espacial. O nervo auditivo só tem trinta mil fibras e transporta mais informação temporal. Estas diferenças pode ser que estejam relacionadas com o que chamamos de *qualia*.

Do que não cabe dúvida alguma é que, se defina como se defina a consciência, a memória e a predição desempenham um papel crucial em sua criação.

As noções de mente e alma estão relacionadas com a consciência.

Quando criança costumava perguntar-me como teria sido se "eu" tivesse nascido no corpo de outra criança em outro país, como se "eu" fosse algo independente do meu corpo. Este sentimento de uma mente independente da fisicalidade é comum e uma consequência natural do funcionamento do neocórtex. Nosso córtex cria um modelo do mundo em sua memória hierárquica. Os pensamentos são o que surge quando este modelo funciona por si mesmo; a lembrança da memória leva a uma nova lembrança da memória, e assim sucessivamente. Nossos pensamentos mais contemplativos não estão dirigidos pelo mundo real, nem sequer estão ligados com ele; são puras criações do nosso modelo. Fechamos os olhos e buscamos em silêncio para que nosso pensamento não seja interrompido pelas entradas sensoriais. Nosso modelo, certamente, criou-se originalmente mediante a exposição ao mundo real através dos sentidos, mas quando planejamos e pensamos sobre ele o fazemos mediante o modelo cortical, não pelo mundo em si.

Para o córtex, nossos corpos não são mais do que parte do mundo exterior. Recordemos que o cérebro é uma caixa silenciosa e escura. Só sabe do mundo através dos padrões das fibras nervosas sensoriais. Desde a perspectiva do cérebro como um aparelho de padrões, ele não sabe do nosso corpo de forma diferente de como sabe do resto do mundo. Não existe uma distinção especial entre onde termina o corpo e onde começa o resto do mundo. Mas o córtex não tem capacidade para modelar o cérebro porque não há sentidos nele. Deste modo, podemos ver por que nossos pensamentos parecem independentes do nosso corpo, por que parece que temos uma mente ou alma independente. O córtex constrói um modelo do nosso corpo, mas não pode construir um modelo do mesmo cérebro. Nossos pensamentos, que se localizam no cérebro, estão separados fisicamente do corpo e do resto do mundo. A mente é independente do corpo, mas não do cérebro.

Podemos observar com clareza esta diferenciação mediante o trauma e a doença. Se alguém perde um membro, o modelo que tem seu cérebro dele pode permanecer intacto, dando como resultado o chamado "membro fantasma" no qual ele é capaz de continuar o sentindo unido ao seu corpo. Por outro lado, se ele sofre um trauma cortical, ele pode perder seu modelo do braço ainda que ele continue conservando-o. Neste caso ele pode padecer da síndrome do braço alheio e ter a sensação incômoda e talvez intolerável de que o braço não é seu e é controlado por outra pessoa. Se nosso cérebro permanece intacto enquanto o resto do corpo cai doente, dá-nos a sensação de que temos uma mente sã agarrada em um corpo agonizante, ainda que na realidade o que temos é um cérebro sã agarrado em um corpo agonizante. É natural imaginar que nossa mente continuará após a morte do nosso corpo, mas quando o cérebro morre, a

mente também morre. Esta verdade acaba sendo evidente se nossos cérebros falham antes que nossos corpos. As pessoas com a doença de Alzheimer ou com um dano cerebral sério perdem a mente mesmo que seu corpo permaneça são.

O que é a Imaginação?

Desde o ponto de vista conceitual, a imaginação é bastante simples. Os padrões fluem à cada área cortical desde nossos sentidos ou áreas inferiores da hierarquia da memória. Cada área cortical cria previsões que se reenviam hierarquia abaixo. Para imaginar algo, basta deixar que nossas previsões deem a volta e se convertam em entradas. Sem fazer nada físico, podemos seguir as consequências das nossas previsões. "Se acontece isto, acontecerá isso e logo aquilo", e assim sucessivamente. Experimentamos isto quando preparamos uma reunião de negócios, jogamos xadrez, organizamos um acontecimento esportivo ou fazemos milhares de outras coisas.

No xadrez imaginamo-nos que movemos o cavalo a certa posição e depois imaginamos como ficará o tabuleiro depois da jogada. Com esta imagem em mente, predizemos o que fará nosso rival e como ficará o tabuleiro após sua jogada. Depois predizemos o que faremos, e assim uma e outra vez. Avançamos pelos passos imaginados e suas consequências. Por fim decidimos, baseando nestas sequências de fatos imaginados, se a jogada inicial era boa ou não. Alguns atletas, como os esquiadores de descida cronometrada, podem melhorar seus resultados se repassam mentalmente o percurso da corrida uma e outra vez em sua cabeça. Ao fechar os olhos e imaginar-se cada um dos giros e dos obstáculos, e inclusive se encontrar no pódio dos ganhadores, aumentam suas possibilidades de sucesso. Imaginar não é mais do que outra palavra para planejar. Aí é onde a capacidade preditiva do nosso córtex vale a pena, pois nos permite saber quais serão as consequências das nossas ações antes que as realizemos.

Imaginar requer um mecanismo neural para converter uma previsão em uma entrada. No capítulo 6 propus que nas células da camada 6 é onde ocorre a previsão precisa. Estas células projetam-se a níveis inferiores da hierarquia, mas também para acima às células de entrada da camada 4. Deste modo, as saídas de uma região podem converter-se em suas próprias entradas. Stephen Grossberg, que há muito tempo idealiza modelos do córtex, chama a este circuito da imaginação de "realimentação dobrada". Se fechamos os olhos e imaginamos um hipopótamo, a área visual do nosso córtex ativa-se, da mesma forma que faria se estivéssemos vendo realmente o animal. Vemos o que imaginamos.

O que é a Realidade?

As pessoas perguntam-me com uma expressão de preocupação e assombro, "Quer dizer que nossos cérebros criam um modelo do mundo? E que o modelo pode ser mais importante que a realidade?"

“Bom, sim; em certa medida, assim seria,” respondo. “Mas não existe o mundo fora da minha cabeça?”

Certamente que sim. As pessoas são reais, minha gata é real, as situações sociais nas quais nos encontramos são reais. Mas nosso entendimento do mundo e nossa resposta a ele se baseiam em predições que provem do nosso modelo interno. Em um dado momento do tempo, só podemos sentir diretamente uma parte diminuta do nosso mundo. Essa ínfima parte dita quais memórias se invocarão, mas ela em si não basta para construir o conjunto da nossa percepção em curso. Por exemplo, agora estou digitando em meu escritório e escuto uma chamada na porta dianteira. Sei que minha mãe tem vindo de visita e imagino que ela está no pé da escada, ainda que não tenha a visto nem a escutado na realidade. Não tinha nada na entrada sensorial unido de forma específica com minha mãe. É meu modelo de memória do mundo que prediz que ela está lá por analogia com a experiência passada. A maioria do que percebemos não chega através dos nossos sentidos, mas do que é gerado por nosso modelo de memória interno.

Portanto, a pergunta “O que é a realidade?” depende em boa medida da precisão com que nosso modelo cortical reflita a real natureza do mundo.

Muitos aspectos do mundo que nos rodeia são tão constantes que quase todos os humanos possuem o mesmo modelo interno deles. Desde criança aprendemos que a luz que cai sobre um objeto redondo produz determinada sombra e que podemos calcular a forma da maioria dos objetos pelas pistas do mundo natural. Aprendemos que se jogávamos uma caneca da nossa cadeira, a gravidade sempre a puxava para o solo. Aprendemos texturas, geometria, cores e os ritmos do dia e da noite. Todos aprendemos de forma sistemática as propriedades físicas simples do mundo.

Mas boa parte do nosso modelo do mundo baseia-se no costume, na cultura e no que nossos pais nos ensinam. Estas partes do nosso modelo são menos constantes e poderiam ser muito diferentes para diferentes pessoas. Uma criança que cresce em um lar cheio de carinho e cuidados com pais que respondem a suas necessidades emocionais é provável que chegue à idade adulta predizendo que o mundo é seguro e amável. As crianças maltratadas por um ou ambos pais têm maiores probabilidades de ver os acontecimentos futuros como algo perigoso e cruel, e achar que não se pode confiar em ninguém por mais bem que os tratem depois. Grande parte da psicologia baseia-se nas consequências da experiência, do apego e da educação durante os primeiros anos, porque é neste período que o cérebro estabelece seu primeiro modelo do mundo.

Nossa cultura molda por completo nosso modelo do mundo. Por exemplo, veja como os asiáticos e os ocidentais percebem o espaço e os objetos de forma diferente. Os asiáticos preocupam-se mais com o espaço entre os objetos, enquanto os ocidentais preocupam-se com os objetos, uma diferença que se

traduz em estéticas e modos de resolver problemas separados. As pesquisas têm demonstrado que algumas culturas, como no caso das tribos do Afeganistão e de algumas comunidades da América do Sul, se constroem sobre os princípios da honra e, como resultado, são mais propensos a aceitar o caráter natural da violência. Os credos religiosos diferentes aprendidos no começo da vida podem levar a modelos completamente diferentes de moralidade, modos de tratar aos homens e as mulheres e inclusive do valor da própria vida. É evidente que estes diferentes modelos do mundo não podem estar completamente corretos em um sentido absoluto e universal, ainda que talvez pareçam corretos a um indivíduo. O raciocínio moral, o bom e o mau, aprende-se.

Nossa cultura (e a experiência familiar) ensina-nos estereótipos, que infelizmente são uma parte inevitável da vida. Ao longo deste livro se poderia substituir memória invariável (ou representação invariável) pela palavra *estereotipo* sem que se alterasse de forma substancial o significado. A predição por analogia é muito semelhante ao julgamento por estereotipo. O estereotipo negativo tem consequências sociais terríveis. Se minha teoria da inteligência está correta, não podemos conseguir fazer com que as pessoas se livrem da sua propensão a pensar em estereótipos, porque são o modo de funcionar do córtex. Formar estereótipos é uma característica inerente do cérebro.

O modo de eliminar o dano causado pelos estereótipos é ensinar a nossos filhos a reconhecer os falsos, a ser empáticos e céticos. Precisamos fomentar estas capacidades de pensamento crítico, além de inculcar-lhes os melhores valores que conhecemos. O ceticismo, o núcleo do método científico, é o único modo que entendemos para separar o fato da ficção.

* * *

Espero ter lhe convencido de que a mente não é mais do que um rótulo do que faz o cérebro. Não é uma coisa separada que manipula ou coexiste com as células no cérebro. Os neurônios não são mais do que células. Não existe uma força mística que faça com que as células nervosas particulares ou os grupos de células nervosas se comportem de maneiras que diferem de como o fariam normalmente. Devido a este fato, podemos agora focar nossa atenção para o modo como poderíamos aplicar a capacidade das células cerebrais para recordar e prever — nosso algoritmo cortical— em silício.

O Futuro da Inteligência

É difícil prever os usos finais de uma nova tecnologia. Como temos visto ao longo deste livro, os cérebros efetuam previsões por analogia ao passado. Portanto, nossa inclinação natural é imaginar que uma nova tecnologia se empregará para fazer o mesmo tipo de coisas que efetuava outra anterior. Imaginamos a utilização de uma nova ferramenta para realizar algo conhecido, só que mais rápido, com maior eficácia ou mais barato.

Abundam os exemplos. As pessoas chamaram a ferrovia de “cavalo de ferro”, e o automóvel, de “carruagem sem cavalos”. Durante décadas, o telefone foi abordado dentro do contexto do telégrafo, algo que só devia se usar para comunicar notícias ou acontecimentos importantes; até a década de 1920 as pessoas não o empregavam de maneira informal. A fotografia foi usada a princípio como uma nova forma de retrato. E os filmes cinematográficos se conceituaram como uma variação das obras de teatro, motivo pelo qual as salas de cinema tiveram cortinas que corriam sobre a tela durante boa parte do século XX.

Não obstante, os usos finais de uma nova tecnologia são com frequência inesperados e de muito maior alcance do que nossa imaginação possa captar a princípio. O telefone tem evoluído a uma rede sem fios de voz e dados que permite a duas pessoas quaisquer do planeta se comunicar entre si estejam aonde estiverem mediante voz, texto e imagens. O transistor foi inventado pela Bell Labs em 1947. De imediato pareceu evidente que o aparelho era um grande avanço, mas as aplicações iniciais não foram mais do que melhoras de outras antigas: os transistores substituíram os tubos com vácuo. Isto levou à fabricação de rádios e computadores menores e mais confiáveis, que foi algo importante e apaixonante em sua época, mas as principais diferenças consistiam no tamanho e confiabilidade das máquinas. As aplicações mais revolucionárias dos transistores não se descobriram tão cedo. Foi necessário um período de inovação gradual antes que alguém pudesse conceber o circuito integrado, o microprocessador, o processador de sinais digitais ou o chip de memória. Mesmo assim, o microprocessador foi desenvolvido pela primeira vez em 1970 tendo em mente as calculadoras de mesa. Mais uma vez, as primeiras aplicações limitaram-se a substituir tecnologias existentes. A calculadora eletrônica substituiu a calculadora de mesa mecânica. Os microprocessadores também foram candidatos claros para substituir os solenoides que então se empregavam em certos tipos de controle industrial, como as mudanças dos semáforos. No entanto, anos depois começou a manifestar-se seu verdadeiro poder. Ninguém dessa época foi capaz de prever o computador pessoal moderno, o telefone móvel, a Internet, o *Global Positioning System* (GPS) ou qualquer outra das peças básicas da tecnologia da informação atual.

Do mesmo modo, seria absurdo pensar que podemos prever as aplicações revolucionárias dos sistemas de memória semelhantes ao cérebro. Espero que estas máquinas melhorem a vida em todos os sentidos. Podemos estar seguros de que assim será. Mas prever o futuro da tecnologia para além de alguns anos é impossível. Para dar-se conta disso basta ler alguns dos absurdos prognósticos que têm realizado os futuristas ao longo dos anos. Na década de 1950 predisse-se que para o ano 2000 todos teríamos reatores atômicos em nossos porões e passaríamos as férias na Lua. Mas enquanto não esqueçamos a lição de moral destes contos, não perdemos nada em especular como serão as máquinas inteligentes. No mínimo, podem-se extrair sobre o futuro algumas conclusões gerais e úteis.

As perguntas acabam sendo intrigantes. Podemos construir máquinas inteligentes e, se assim for, que aparência terão? Se aproximarão aos robôs parecidos com os humanos da ficção popular, ao gabinete preto ou bege de um computador pessoal, ou a alguma outra coisa? Como se usarão? Trata-se de uma tecnologia perigosa que possa ferir-nos ou ameaçar nossas liberdades pessoais? Quais são as aplicações óbvias das máquinas inteligentes e há algum modo de saber quais serão as aplicações fantásticas? Qual será a repercussão final das máquinas inteligentes sobre nossas vidas?

Podemos Construir Máquinas Inteligentes?

Sim podemos, mas talvez não sejam o que esperamos. Ainda que talvez pareça óbvio, não acho que vamos construir máquinas inteligentes que atuem como seres humanos, nem interajam conosco de formas humanas.

Uma noção popular das máquinas inteligentes chega-nos dos filmes e livros: são os robôs humanoides amáveis, maus ou às vezes desajeitados que conversam conosco sobre sentimentos, ideias e acontecimentos, e desempenham um papel em inumeráveis enredos fantásticos. Em um século de ficção científica tem-se treinado às pessoas para que considere os robôs e andróides uma parte inevitável e desejável do nosso futuro. Gerações têm crescido com as imagens de Robbie, o robô de *Planeta Proibido*; R2D2 e C3PO, de *Guerra nas Estrelas*, e do tenente comandante Data, de *Star Trek*. Inclusive HAL, do filme *2001: Uma Odisseia no Espaço*, ainda que carente de corpo, era muito parecido com os humanos e tinha sido projetado desta forma para ser um colega como um copiloto programado em sua longa viagem espacial. Os robôs de aplicações limitadas tais como carros inteligentes, minissubmarinos autônomos para explorar as profundidades oceânicas e aspiradores ou cortadores de grama auto-dirigidos— são factíveis, e é bem possível que em algum dia sejam de uso corrente. Mas andróides e robôs como o tenente comandante Data e C3PO vão continuar sendo ficção durante muito tempo. Há um par de razões para isso.

A primeira é que a mente humana não é criada só pelo neocórtex, mas também pelos sistemas emocionais do cérebro velho e pela complexidade do corpo humano. Para ser humano precisa-se de toda a maquinaria biológica, não só o córtex. Para conversar como um ser humano sobre todos os temas (para passar no teste de Turing) se requereria uma máquina inteligente que tivesse a maioria das experiências e emoções de um humano real e vivesse uma vida semelhante à humana. As máquinas inteligentes contarão com o equivalente de um córtex e um conjunto de sentidos, mas o resto é optativo. Talvez pareça divertido observar a uma máquina inteligente se deslocar por aí em corpos parecidos aos humanos, mas possuirão mentes que serão remotamente semelhantes às humanas a não ser que se lhes incorporem sistemas emocionais e experiências semelhantes às nossas, o que seria difícilíssimo e, ao meu parecer, careceria de sentido.

A segunda razão é que, dado o custo e o esforço necessários para construir e manter robôs humanoides, custa acreditar que acabassem sendo práticos. Um mordomo robô seria mais caro e menos útil do que um assistente humano. Ainda que o robô fosse "inteligente", não teria a relação e entendimento fácil que mostra um assistente humano pelo simples fato de ser um semelhante humano.

Tanto a máquina à vapor como o computador digital suscitaram visões robóticas que nunca chegaram a dar frutos. Da mesma forma, quando pensamos em construir máquinas inteligentes, para muitas pessoas lhes acaba sendo natural imaginar robôs humanoides, mas não é provável que isto se torne realidade. Os robôs são um conceito nascido da Revolução Industrial e aperfeiçoado pela ficção. Não devemos ir a eles em busca de inspiração para desenvolver máquinas realmente inteligentes.

Portanto, que aparência terão as máquinas inteligentes se não são robôs que caminham e falam? A evolução descobriu que se unisse a nossos sentidos um sistema de memória hierárquica, a memória modelaria o mundo e iria predizer o futuro. Copiando a Natureza, devemos construir máquinas inteligentes com os mesmos princípios. Aí está a receita para construir as ditas máquinas. Começa-se com um conjunto de sentidos para extrair padrões do mundo. É provável que nossa máquina inteligente tenha um conjunto de sentidos diferentes dos humanos e inclusive talvez "exista" em um mundo diferente do nosso. Portanto, não há por que presumir que ela tem de ter um jogo de globos oculares e um par de orelhas. A seguir deve-se acrescentar a esses sentidos um sistema de memória hierárquica que funcione segundo os mesmos princípios que o córtex. Depois teremos que treinar a esse sistema de memória de forma muito semelhante a como se ensina as crianças. Depois de sessões de treinamento repetitivo, nossa máquina inteligente construirá um modelo do seu mundo tal como o veem *seus* sentidos. Não terá necessidade nem oportunidade para que ninguém lhe programe as regras do mundo, bases de dados, fatos ou quaisquer dos conceitos de alto nível que são o pesadelo da inteligência artificial. A máquina inteligente deve aprender através

da observação do seu mundo e incluir as entradas de um instrutor quando for preciso. Uma vez que nossa máquina inteligente tem criado um modelo do seu mundo, ela pode descobrir analogias com experiências passadas, efetuar previsões sobre acontecimentos futuros, propor soluções a novos problemas e pôr a nossa disposição o dito conhecimento.

Desde o ponto de vista físico, nossa máquina inteligente pode incorporar-se a aviões ou carros, ou permanecer estoicamente em uma estante de uma sala de computadores. Diferente dos seres humanos, cujos cérebros devem acompanhar a seus corpos, o sistema de memória de uma máquina inteligente poderia localizar-se bem longe dos seus sensores (e "corpo", se o tivesse). Por exemplo, um sistema de segurança inteligente poderia ter sensores colocados por toda uma fábrica ou cidade, mas o sistema de memória hierárquica unido a esses sensores estaria encerrado no porão de um edifício. Portanto, a forma física de uma máquina inteligente teria a possibilidade de adotar muitas formas.

Não há razão para que uma máquina inteligente apresente a aparência de um ser humano e atue ou sinta como ele. O que a faz inteligente é que ela compreende e interage com seu mundo valendo de um modelo de memória hierárquica e é capaz de pensar sobre seu mundo de modo análogo a como você e eu pensamos sobre o nosso. Como veremos, seus pensamentos e ações poderiam ser completamente diferentes dos de um ser humano, mas continuaria sendo inteligente. A inteligência mede-se pela capacidade preditiva de uma memória hierárquica, não por um comportamento semelhante ao humano.

* * *

Prestemos atenção agora ao maior desafio que enfrentaremos quando construirmos máquinas inteligentes: a criação da memória. Para conseguir isto, precisaremos compor grandes sistemas de memória organizados hierarquicamente e que funcionem como o córtex. A capacidade e a conectividade significarão grandes avanços.

A capacidade é o primeiro assunto. Digamos que o córtex conte com trinta e dois trilhões de sinapses. Se assumimos que cada uma utiliza só dois bits (que nos dão quatro possíveis valores por sinapses) e que cada byte tem oito bits (portanto, um byte poderia representar quatro sinapses), precisaríamos de uns oito trilhões de bytes de memória. Um disco rígido de um computador pessoal atual possui cerca de cem bilhões de bytes, de modo que precisaríamos de uns oitenta discos rígidos atuais para contar com a mesma quantidade de memória que o córtex. (Não se preocupe com as cifras exatas, pois não são mais do que cálculos aproximados.) O importante é que esta quantidade de memória seja construível no laboratório. Por mais que erremos nos cálculos, não é o tipo de máquina que cabe em um bolso ou que se possa incorporar em uma torradeira. De todos modos, o importante é que a quantidade de memória requerida não se encontra fora do nosso alcance,

enquanto que há só dez anos atrás a teria estado. Ajuda-nos o fato de que não temos que recriar um córtex humano inteiro, pois muito menos bastará para diversas aplicações.

Nossas máquinas inteligentes precisarão de muita memória. É provável que comecemos a construir utilizando discos rígidos ou discos ópticos, mas acabaremos as fabricando também de silício. Os chips de silício são pequenos, gastam pouca energia e são resistentes. E é só questão de tempo para que se possam fabricar chips de memória de silício com a capacidade precisa para construir máquinas inteligentes. De fato, a memória inteligente tem uma vantagem sobre a memória computacional convencional. A economia da indústria dos semicondutores baseia-se no percentual de chips com erros. O percentual de chips bons recebe o nome de colheita e determina se o projeto de um chip particular pode ser fabricado e vendido lucrativamente. Como a possibilidade de erro aumenta à medida que é feito o chip, a maioria dos atuais não excedem o tamanho de um selo postal. A indústria não tem potencializado a quantidade de memória em um único chip o fazendo maior, mas, em linhas gerais, reduzindo suas características particulares.

Mas os chips inteligentes de memória têm que tolerar as falhas. Recordemos que nenhum componente único dos nossos cérebros contém um elemento indispensável de dados. Nossos cérebros perdem milhares de neurônios cada dia, mas nossa capacidade mental vai decrescendo a ritmo lento ao longo da vida adulta. Os chips de memória inteligentes funcionarão segundo os mesmos princípios que o córtex, de modo que ainda que um percentual dos elementos da memória acabem sendo defeituosos, o chip continuará sendo útil e rentável comercialmente. É muito provável que a tolerância aos erros da memória semelhante ao cérebro permita aos projetistas construir chips que sejam bem mais grandes e densos que os chips de memória dos computadores atuais. O resultado é que talvez contemos com um cérebro fabricado em silício antes do que as tendências presentes indicam.

O segundo problema que devemos superar é a *conectividade*. Os cérebros reais possuem grandes quantidades de matéria branca subcortical. Como já temos notado, a matéria branca está composta por milhões de axônios que fluem de um lado a outro sob a fina lâmina cortical e ligam entre si as diferentes regiões da hierarquia cortical. Uma célula particular do córtex pode ligar-se com outras cinco ou dez mil células. Este tipo de conexão paralela em massa é difícil ou impossível de conseguir utilizando as técnicas de fabricação de silício tradicionais. Os chips de silício fabricam-se depositando algumas camadas de metal, cada uma separada por uma camada de isolamento. (Este processo não tem nada a ver com as camadas do córtex.) As camadas de metal contêm os "cabos" do chip, e como não podem atravessar a que lhes corresponde, o número total de conexões é limitado. Não vai servir para os sistemas de memória semelhantes ao cérebro, nos quais se precisam

de milhões de conexões. Os chips de silício e a matéria branca não são muito compatíveis.

Será preciso de muita engenharia e experimentação para resolver este problema, mas sabemos o básico sobre o modo de solucionar isto. Os cabos elétricos enviam sinais bem mais rápido que os axônios dos neurônios. Um único cabo de um chip pode ser compartilhado e, portanto, ser utilizado para muitas conexões diferentes, enquanto que no cérebro cada axônio pertence a um único neurônio.

Um exemplo do mundo real é o sistema telefônico. Se tecessemos uma linha de cada telefone a outro, a superfície do globo ficaria enterrada sob uma selva de alambre de cobre. O que fazemos ao invés disso é viabilizar que todos os telefones compartilhem um número relativamente pequeno de linhas de alta capacidade. Este método funciona sempre que a capacidade de cada linha é muito maior que a requerida para transmitir uma única conversa. O sistema telefônico cumpre este requisito: um único cabo de fibra óptica pode transportar ao mesmo tempo um milhão de conversas.

Os cérebros reais possuem axônios exclusivos entre todas as células que falam entre si, mas podemos construir máquinas inteligentes que se assemelhem mais ao sistema telefônico e que compartilhem conexões. Acredite ou não, alguns cientistas vêm pensando em como resolver o problema da conectividade do chip cerebral durante muitos anos. Ainda que o funcionamento do córtex continuasse sendo um mistério, os pesquisadores sabiam que em algum dia desvendaríamos o quebra-cabeça e então teríamos que abordar o tema da conectividade. Não é preciso que analisemos aqui as diferentes propostas. Basta dizer que a conectividade poderia ser o maior obstáculo técnico que nós encontraríamos para construir máquinas inteligentes, mas temos que ser capazes de resolver isto.

Uma vez que se superarem os desafios tecnológicos, não há problemas fundamentais que nos impeça de construir sistemas realmente inteligentes. É verdade que existe uma multidão de temas que terão que ser abordados para que estes sistemas sejam pequenos, de baixo custo e gastem pouca energia, mas nada se interpõe em nosso caminho. Precisaram-se de cinquenta anos para passar dos computadores do tamanho de uma sala para os que cabem no bolso, mas como partimos de uma posição tecnológica avançada, a transição para as máquinas inteligentes tem que ser bem mais rápida.

Devemos Construir Máquinas Inteligentes?

No século XXI as máquinas inteligentes sairão do reino da ficção científica para se converterem em realidade. Antes que chegue esse momento, devemos pensar sobre assuntos éticos, se os perigos possíveis superarão os benefícios prováveis.

A perspectiva de contar com máquinas capazes de pensar e agir por conta própria tem preocupado as pessoas há muito tempo, e é compreensível. Os novos campos do conhecimento e as tecnologias inovadoras sempre amedrontam quando aparecem. A criatividade humana faz-nos imaginar formas terríveis nas quais uma nova tecnologia pode apoderar dos nossos corpos, nos tornar inúteis ou suprimir o próprio valor da vida humana. Mas a história mostra que essas obscuras imaginações quase nunca se cumprem do modo que esperamos. Quando se produziu a Revolução Industrial, ficamos com medo da eletricidade (recordam de Frankenstein?) e dos motores à vapor. A maquinaria que contava com energia própria, que podia se mover de formas complexas, parecia milagrosa e ao mesmo tempo potencialmente sinistra. Mas a eletricidade e os motores de combustão interno já não nos parecem estranhos nem sinistros. Fazem parte do nosso ambiente, da mesma forma que o ar e a água.

Quando começou a revolução da informática, em seguida começamos a temer os computadores. Tinha inúmeras histórias de ficção científica sobre computadores potentes ou redes computacionais que de forma espontânea ganhavam consciência de si mesmos e depois se faziam de amos. Mas agora que os computadores se integraram na vida diária, esse temor parece absurdo. O computador da sua casa ou a Internet têm as mesmas possibilidades de ganhar consciência que uma caixa registradora.

Toda tecnologia pode se aplicar para fins bons ou maus, certamente, mas algumas são mais propensas ao mau uso ou a catástrofe que outras. A energia atômica é perigosa tanto na forma de ogivas nucleares como na de centrais nucleares, porque um único acidente ou um único mau uso poderia ferir ou matar a milhões de pessoas. E ainda que a energia nuclear seja valiosa, dispõe-se de alternativas. A tecnologia veicular pode adotar a forma de tanques e caças de combate, ou a de carros e aviões de passageiros, e um percalço ou mau uso pode causar dano a muita gente. Mas caberia afirmar que os veículos são mais essenciais para a vida moderna e menos perigosos que a energia nuclear. O dano causado por um único mau uso de um avião é muito menor que o de uma bomba nuclear. Existem muitas tecnologias que são quase que completamente benéficas. Os telefones são um exemplo. Sua tendência para comunicar e unir as pessoas ultrapassa significativamente qualquer efeito negativo. O mesmo é aplicável à eletricidade e à ciência da saúde pública. Na minha opinião, as máquinas inteligentes vão ser uma das tecnologias menos perigosas e mais benéficas que jamais tenhamos desenvolvido.

No entanto, alguns pensadores, como Bill Joy, cofundador da Sun Microsystems, temem que desenvolvamos nano-robôs inteligentes que escapem do nosso controle, invadam a Terra e a refaçam, átomo por átomo, segundo seus próprios interesses. A imagem traz-me à mente essas vassouras animadas por magia do aprendiz de bruxo, que se regeneram dos seus estilhaços e trabalham incansáveis

para provocar o desastre. Em linhas similares, alguns otimistas da inteligência artificial oferecem profecias de extensão da vida que parecem inquietantes. Por exemplo, Ray Kurzweil fala do dia em que os nano-robôs serão introduzidos no nosso cérebro para gravar cada sinapse e cada conexão e depois passar toda a informação a um supercomputador, que se reconfigurará em nós. Nos converteremos em uma versão de software de nós mesmos que será praticamente imortal. Estes dois temores sobre a inteligência das máquinas, o cenário das máquinas inteligentes que fazem estragos e o cenário da descarga dos nossos cérebros em um computador, parecem voltar à tona uma e outra vez.

Construir máquinas inteligentes não é o mesmo que construir máquinas auto-reprodutoras. Não há nenhuma conexão lógica entre ambas coisas. Nem os cérebros nem os computadores podem duplicar-se de forma direta, e os sistemas de memória semelhantes aos cérebros não serão diferentes. Ainda que um dos pontos fortes das máquinas inteligentes será nossa capacidade para as produzir em série, é algo muito diferente da duplicação das bactérias e dos vírus. A duplicação não requer inteligência, e a inteligência não requer duplicação.

Mesmo assim, duvido seriamente de que consigamos copiar nossas mentes em máquinas. Na atualidade, até onde sei, não existem métodos reais ou imaginados capazes de gravar os trilhões de detalhes que criam "você". Precisariamos gravar e recriar todo nosso sistema nervoso e nosso corpo, não só o neocórtex. E seria preciso compreender como funciona tudo isso. Sem dúvida, algum dia talvez sejamos capazes de conseguir, mas os desafios se estendem para bem mais além do entendimento do funcionamento do córtex. Imaginar seu algoritmo e incorporar às máquinas do zero é uma coisa, mas escanear os trocentos detalhes operativos de um cérebro vivo e duplicar em uma máquina é algo completamente diferente.

* * *

Mas além da duplicação e da cópia de mentes, as pessoas se preocupam com outra coisa das máquinas inteligentes. Poderiam ameaçar de algum modo a grandes quantidades de população, como fazem as bombas nucleares? Poderia sua presença fazer com que pequenos grupos ou indivíduos malévolos se tornassem muito poderosos? Ou poderiam tornar-se más e agir contra nós, como as máquinas implacáveis dos filmes *Exterminador do Futuro* e *Matrix*?

A resposta a estas perguntas é não. Como aparelhos de informação, os sistemas de memória semelhantes aos cérebros se encontrarão entre as tecnologias mais úteis que temos desenvolvido até o momento. Mas, assim como os carros e os computadores, só serão ferramentas. Que vão ser inteligentes não significa que terão uma habilidade especial para destruir a propriedade ou manipular as pessoas. E do mesmo modo que não poríamos o controle do arsenal nuclear do mundo sob a autoridade de uma pessoa ou um computador, teremos de ter

cuidado para não depender muito delas, pois falharão como qualquer outra tecnologia.

Isto nos leva à questão da malevolência. Algumas pessoas presumem que ser inteligente equivale basicamente a ter mentalidade humana. Temem que as máquinas inteligentes se ressentam de estar “escravizadas”, porque os humanos detestam essa condição. Temem que as máquinas inteligentes tentem se apoderar do mundo, porque as pessoas inteligentes ao longo da história tem tratado de o dominar. Mas estes temores descansam em uma analogia falsa. Baseiam-se em uma refundição da inteligência —o algoritmo neocortical— com os impulsos emocionais do cérebro velho —coisas como temor, paranoia e desejo—. Mas as máquinas inteligentes não terão estas faculdades. Não terão ambição pessoal. Não desejarão riqueza, reconhecimento social ou gratificação sensual. Não terão apetites, vícios ou transtornos de caráter. As máquinas inteligentes não possuirão nada que se pareça à emoção humana, a não ser que nós tomemos o trabalho de as projetar para tal efeito. Suas aplicações mais valiosas serão onde o intelecto humano encontra dificuldade, áreas nas quais nossos sentidos acabam sendo inadequados, ou atividades que nos acabam sendo tediosas. Em geral, estas atividades possuem pouco conteúdo emocional.

As máquinas inteligentes abrangerão desde sistemas simples de uma única aplicação até sistemas inteligentes sobre-humanos muito potentes; mas a não ser que desviemo-nos do caminho para fazê-las semelhantes aos humanos, não o serão. Talvez em algum dia tenhamos que pôr restrições ao que se pode fazer com elas, mas é algo que está muito longe e, quando chegar, é provável que os assuntos éticos acabem sendo relativamente simples comparados com questões morais atuais como as que rodeiam a genética e a tecnologia nuclear.

Por que Construir Máquinas Inteligentes?

Analisemos agora o que farão as máquinas inteligentes.

Pedem-me com frequência para que eu dê palestras sobre o futuro da computação móvel. O organizador de uma conferência me solicita que eu descreva como serão os computadores de bolso ou os telefones móveis em cinco ou vinte anos. Querem escutar minha visão do futuro, mas não posso a dar. Para estabelecer minha premissa, uma vez saí em cena com um chapéu de mago e uma bola de cristal. Expliquei que ninguém é capaz de ver o futuro com detalhe. As bolas de cristal são ficção, e qualquer pessoa que pretenda saber com exatidão o que acontecerá nos anos futuros falhará com total segurança. O melhor que se tem a fazer é compreender tendências gerais. Se entendemos uma ideia geral, seremos capazes de segui-la onde quer que vá enquanto se desdobram os detalhes.

O mais famoso exemplo da tendência de uma tecnologia é a lei de Moore. Gordon Moore acertou ao predizer que o número de elementos de circuito que poderiam

colocar em um chip de silício se duplicaria cada ano e meio. Ele não disse se seriam chips de memória, unidades de processamento central ou outra coisa. Ele não indicou para que tipo de produtos se utilizariam os chips. Ele não predisse se os chips se alojariam em plástico ou cerâmica, ou se seriam colocados em placas de circuito. Ele não disse nada sobre os diversos processos empregados para fabricar os chips. Ele limitou-se a expressar a tendência mais ampla que ele pôde, e acertou.

Hoje, por si só, não podemos predizer os usos finais das máquinas inteligentes, pois não há modo de acertar os detalhes. Se eu ou outra pessoa predizemos minuciosamente o que farão essas máquinas, será inevitável que nos equivoquemos. Não obstante, pode-se fazer um pouco mais do que nos limitar a nos encolher de ombros. Existem duas linhas de pensamento úteis. Uma é conjecturar os usos a muito curto prazo dos sistemas de memória semelhantes ao cérebro: as coisas óbvias mas menos interessantes que se tentarão primeiro. A segunda proposta é pensar sobre as tendências à longo prazo, como a lei de Moore, que podem nos ajudar a imaginar as aplicações que provavelmente farão parte do nosso futuro.

Começamos com algumas aplicações à curto prazo. São as coisas que parecem óbvias, como substituir os tubos de um rádio por transistores ou construir calculadoras com um microprocessador. E podemos começar observando algumas áreas que a inteligência artificial tentou desvendar mas não pôde: reconhecimento de voz, visão e carros inteligentes.

* * *

Se alguma vez você tentou usar *software* de reconhecimento de voz para introduzir texto em um computador pessoal, sabe o absurdo que isto pode resultar. Assim como a sala chinesa de Searle, o computador não tem conhecimento do que se lhe diz. Nas poucas vezes que provei esses produtos fiquei frustrado. Se houvesse algum ruído na sala, desde um lápis que caísse até alguém que me falasse, apareciam palavras adicionais na minha tela. Os índices de erro eram elevados. Com frequência as palavras que o *software* pensava que eu dizia careciam de sentido. "Lembre de dizer a Mary que o chão está preparado para passear." Uma criança daria conta de que era um erro, mas não um computador. De forma similar, as chamadas interfaces de linguagem natural têm sido uma meta para os cientistas da computação durante anos. A ideia é de que sejamos capazes de dizer a um computador ou a outro eletrodoméstico o que queremos em linguagem simples e deixar que a máquina realize o trabalho. A um assistente digital pessoal ou PDA se lhe poderia indicar: "Transfira a partida de basquete da minha filha para o domingo às dez da manhã". Este tipo de coisas tem sido impossível fazê-las bem com a inteligência artificial tradicional. Ainda que o computador pudesse reconhecer cada palavra, para completar a tarefa precisaria saber onde fica o colégio da sua filha, que provavelmente você quer dizer no

próximo domingo e talvez que seja uma partida de basquete, porque o compromisso só poderia dizer "Menlo contra St. Joe". Ou talvez queiramos que um computador escute uma transmissão de rádio em busca da menção de um produto particular, mas o locutor descreve o produto sem utilizar seu nome. Você e eu saberíamos do que ele está falando, mas não um computador.

Estas e outras aplicações requerem que a máquina seja capaz de escutar a linguagem falada. Mas os computadores não conseguem realizar estas tarefas porque não entendem o que se diz. Casam padrões auditivos com molde de palavras pelo tom, sem saber o que significam as palavras. Imaginemos que aprendêssemos a reconhecer os sons das palavras particulares de uma língua estrangeira, mas não o seu significado, e nos pedissem que transcrevêssemos uma conversa na dita língua. A conversa vai discorrendo e não temos ideias do que trata, mas tentamos recolher as palavras isoladas. No entanto, as palavras se sobrepõem e se interferem, e perdem-se partes do som pelo ruído. Nos acabaria sendo extremamente difícil separar palavras e as reconhecer. O *software* de reconhecimento de voz luta contra estes obstáculos na atualidade. Os engenheiros têm descoberto que utilizando probabilidades de transições de palavras podem melhorar um pouco sua precisão. Por exemplo, eles empregam regras gramaticais para decidir entre dois homônimos. É uma forma muito simples de predição, mas os sistemas continuam parecendo absurdos. Só conseguem sucesso em situações muito reduzidas nas quais o número de palavras que se poderiam pronunciar em um dado momento seja limitado. No entanto, os seres humanos realizam com facilidade muitas tarefas relacionadas com a linguagem porque nosso córtex entende não só as palavras, mas as orações e o contexto no qual se expressam. Adiantamos ideias, expressões e palavras particulares. Nosso modelo cortical do mundo faz isto de forma automática.

Portanto, cabe esperar que os sistemas de memória semelhantes ao córtex transformem o reconhecimento de voz falível em um entendimento ajustado da fala. Em vez de programar probabilidades para transições de uma única palavra, a memória hierárquica rastreará acentos, palavras, expressões e ideias, e os utilizará para interpretar o que se está dizendo. Assim como uma pessoa, essa máquina inteligente poderia distinguir entre vários fatos da fala; por exemplo, uma discussão entre você e um amigo na sala, uma conversa telefônica e os comandos de edição para um livro. Não será fácil construir as ditas máquinas. Para entender por completo a linguagem humana, a máquina terá que experimentar e aprender o que os humanos fazem. Portanto, ainda que talvez demoremos alguns anos para conseguir construir uma máquina inteligente que entenda a linguagem tão bem como você e eu, em curto prazo seremos capazes de melhorar os resultados dos sistemas de reconhecimento de voz existentes empregando memórias semelhantes ao córtex.

A visão oferece outro conjunto de aplicações que a inteligência artificial tem sido incapaz de conseguir mas que os sistemas inteligentes reais devem manejar. Na atualidade não há uma máquina que possa olhar uma cena natural —o mundo diante dos nossos olhos— ou uma foto usando uma câmera e descrever o que vê. Existem algumas aplicações de visão em máquinas que funcionam em campos muito restringidos, como a alinação visual de chips em uma placa de circuito ou a correlação de características faciais com bases de dados, mas é impossível para um computador identificar uma variedade de objetos ou analisar uma cena de modo mais geral. Não nos custa nada olhar ao redor de uma sala e encontrar um lugar aonde nos sentar, mas não pedamos a um computador que faça isto. Imaginemos que olhamos a tela de vídeo de uma câmera de segurança. Poderíamos notar a diferença entre alguém que chama à porta segurando um presente e alguém que golpeia a porta com um pé-de-cabra? Certamente que sim, mas a distinção está bem além da capacidade do *software* atual. Em consequência, contratamos pessoas para que fiquem de olho nas telas das câmeras de segurança vinte e quatro horas em busca de algo suspeito. Para vigilantes humanos acabam sendo difícil permanecer alerta, enquanto uma máquina inteligente poderia realizar a tarefa sem cansar-se. Muitas situações que dependem da acuracidade visual requerem, além disso, o entendimento de uma cena complexa. As máquinas inteligentes são o único modo de atendê-las.

Por último, analisemos o transporte. Os carros estão ficando muito sofisticados. Eles possuem GPS para traçar a rota de A a B, sensores para acender as luzes quando anoitece, acelerômetros para ativar os *airbags* e sensores de proximidade para nos indicar que estamos a ponto de colidir com algo. Inclusive há carros que podem se conduzir de forma autônoma em autopistas especiais ou em condições ideais, ainda que não sejam comercializados. Mas para conduzir um carro com segurança e eficácia em todo tipo de estradas e condições de tráfego se requer mais do que alguns sensores e circuitos de controle de realimentação. Para ser um bom motorista deve-se compreender o tráfego, os outros motoristas, o funcionamento dos carros, os semáforos e muitas outras coisas. É preciso ser capaz de entender os sinais de perigo ou dar-se conta de que outro carro está sendo conduzido de forma perigosa. É necessário ver o pisca-alerta de outro carro e prever que provavelmente ele mudará de pista, ou se a luz permanecer acesa por vários minutos, discernir que provavelmente o motorista não tenha se dado conta e, portanto, não mudará de pista. É preciso reconhecer que uma coluna de fumaça lá na frente poderia significar que ocorreu um acidente e, portanto, deve-se reduzir a velocidade. Um motorista que vê uma bola cruzar a rua, e pensa de forma automática que uma criança correrá para apanhar, freia por intuição.

Digamos que queremos construir um carro realmente inteligente. A primeira coisa que faríamos é selecionar um conjunto de sensores que lhe permitissem sentir seu mundo. Poderíamos começar com uma câmera para ver, talvez câmeras múltiplas na frente e atrás, e microfones para ouvir, mas talvez também quereríamos lhe pôr

radar ou sensores de ultrassom capazes de determinar com precisão o alcance e a velocidade de outros objetos em condições de escassa visibilidade. O que quero assinalar é que não temos que recorrer aos sentidos que usam os seres humanos, nem nos restringir a eles. O algoritmo cortical é flexível e, sempre que projetarmos nosso sistema de memória hierárquica de forma adequada, ele funcionará independentemente do tipo de sensores que instalarmos. Em teoria, nosso carro poderia ultrapassar-nos na percepção do mundo do tráfico porque seu conjunto de sentidos foi escolhido para cumprir essa tarefa. Os sensores se acoplarão a um sistema de memória hierárquica em uma amplitude suficiente. Os projetistas do carro treinarão sua memória expondo às condições do mundo real para que ele aprenda a construir um modelo do seu mundo do mesmo modo que o fazem os humanos, só que em um campo mais limitado. (Por exemplo, o carro precisa saber de estradas, mas não de elevadores e aviões.) Sua memória aprenderia a estrutura hierárquica do tráfico e das estradas para que entenda e preveja o que é provável que aconteça em seu mundo de automóveis em movimento, sinais de tráfico, obstáculos e intersecções. Os engenheiros poderiam projetar o sistema de memória para que conduza realmente o carro ou se limite a explorar o que acontece quando nós conduzimos. Poderia dar conselhos ou assumir o controle em situações extremas, como se fosse um motorista do assento traseiro do qual não levamos a mal que o fizesse. Uma vez que a memória estivesse bem treinada e o carro fosse capaz de compreender e resolver o que pudesse acontecer, os engenheiros teriam a opção de a configurar de forma permanente para que todos os carros que saíssem da linha de montagem se comportassem da mesma forma, ou poderiam a projetar para que continuasse aprendendo uma vez que fosse vendido o carro. E como no caso de um computador, mas não de um ser humano, a memória poderia ser reprogramada com uma versão atualizada se as novas condições a requeressem.

Não digo que vamos construir carros inteligentes ou máquinas que compreendam a linguagem e a visão, mas são bons exemplos do tipo de aparelhos que poderíamos pesquisar e desenvolver, e que parece possível construir.

* * *

Pelo o que se refere a mim, sinto pouco interesse pelas aplicações óbvias das máquinas inteligentes. Acho que o benefício e estímulo verdadeiros de uma nova tecnologia é encontrar-lhe usos que antes pareciam inconcebíveis. De que formas nos surpreenderão as máquinas inteligentes e que capacidades fantásticas surgirão com o tempo? Estou seguro de que as memórias hierárquicas, assim como o transistor e o microprocessador, transformarão nossas vidas para melhor de modos incríveis; mas como? Uma maneira de vislumbrar o futuro das máquinas inteligentes é pensar em aspectos da tecnologia que sejam escaláveis; isto é, que atributos das máquinas inteligentes irão se tornar cada vez mais baratos, cada vez mais rápidos ou cada vez mais pequenos. As coisas que crescem a taxas

exponenciais logo ultrapassam a nossa imaginação e é muito provável que desempenhem um papel chave nas evoluções mais radicais da tecnologia futura.

Entre os exemplos de tecnologias que têm melhorado de forma exponencial durante muitos anos se inclui o chip de memória de silício, o disco rígido, as técnicas de sequenciação do DNA e a transmissão por fibra óptica. Estas tecnologias têm sido a base de muitos produtos e empresas novos. De um modo diferente, o *software* também é escalável. Um programa desejado, uma vez escrito, pode copiar-se infinitamente quase sem custo algum.

Em contraste, algumas tecnologias, como as baterias, os motores e a robótica tradicional, não são escaláveis. Apesar da multidão de esforços e melhoras constantes, um braço robótico construído hoje não é muito melhor que outro construído há dez anos. Os avanços da robótica são graduais e modestos, e acham-se bem longe das curvas de crescimento exponencial do projeto de chips ou da proliferação de software. Um braço de robô construído em 1985 por um milhão de dólares não se pode construir hoje mil vezes mais forte por apenas dez dólares. Do mesmo modo, as baterias de hoje não são muitíssimo melhores que as de dez anos atrás. Caberia afirmar que são duas ou três vezes melhores, mas não mil ou dez mil vezes melhores, e o progresso avança muito lento. Se a capacidade das baterias aumentasse no mesmo ritmo que a dos discos rígidos, dos telefones móveis e de outros aparelhos eletrônicos, jamais teriam que recargar-se e os carros elétricos rápidos que percorressem 1.600 quilômetros por carga se tornariam comuns.

Portanto, isto faz-nos pensar que aspectos dos sistemas de memória semelhante ao cérebro acabarão ultrapassando de forma espetacular os nossos cérebros biológicos. Estes atributos nos sugerirão onde terminará a tecnologia. Vejo quatro atributos que superarão nossas faculdades: velocidade, capacidade, possibilidade de duplicação e sistemas sensoriais.

Velocidade

Enquanto os neurônios funcionam na ordem de milissegundos, o silício funciona na de nanosegundos (e continua ganhando velocidade), o que significa uma diferença de um milhão de vezes ou seis ordens de magnitude. A diferença de velocidade entre as mentes orgânicas e as baseadas em silício será de grande importância. As máquinas inteligentes serão capazes de pensar um milhão de vezes mais rápido do que o cérebro humano. A dita mente poderia ler bibliotecas inteiras ou estudar enormes compilações de dados complexos em poucos minutos, extraindo o mesmo entendimento. Não há nada de magia nisso. Os cérebros biológicos evoluíram com duas limitações relacionadas com o tempo. Uma é a velocidade com que as células podem fazer as coisas e a outra é a velocidade em que muda o mundo. Talvez não seja muito útil para um cérebro biológico pensar um milhão de vezes mais rápido se o mundo que o rodeia é lento por si só. Mas não há nada no algoritmo cortical

que diga que ele deva operar sempre devagar. Se uma máquina inteligente conversasse ou interagisse com um humano, ela teria que diminuir a marcha para funcionar na velocidade humana. Se lesse um livro passando páginas, ela teria que ter um limite na velocidade com que poderia o fazer. Mas quando ela se ligasse com o mundo eletrônico, poderia funcionar bem mais rápido. Duas máquinas inteligentes poderiam manter uma conversa um milhão de vezes mais rápida que dois humanos. Imaginemos o avanço de uma máquina inteligente que resolvesse problemas matemáticos ou científicos um milhão de vezes mais rápido que um ser humano. Em dez segundos poderia dedicar mais reflexão a um problema do que nós conseguiríamos em um mês. Sem dúvida, mentes que não se cansam nem se entediam jamais com essa velocidade exorbitante acabarão sendo úteis de formas que ainda não podemos imaginar.

Capacidade

Apesar da impressionante capacidade de memória de um córtex humano, as máquinas inteligentes que construirmos poderão a ultrapassar significativamente. O tamanho do nosso cérebro viu-se limitado por vários fatores biológicos, entre os quais se incluem a proporção do crânio da criança em relação ao diâmetro pélvico da mãe, o alto custo metabólico do funcionamento cerebral (nosso cérebro possui em torno de 2 por cento do peso corporal, mas usa quase 20 por cento do oxigênio que respiramos), e a lentidão dos neurônios. Mas podemos construir sistemas de memória inteligentes de qualquer tamanho, e incluir previsão e uma intenção específica aos detalhes do projeto. A capacidade do neocórtex humano talvez acabe sendo relativamente modesta dentro de umas décadas.

Ao construir máquinas inteligentes, poderíamos aumentar sua capacidade de memória de vários modos. Se fosse acrescentada profundidade à hierarquia, se conseguiria um entendimento mais penetrante, pois se veriam padrões de ordens superiores. Se fosse ampliada a capacidade das regiões, a máquina recordaria mais detalhes, ou perceberia com maior agudeza, do mesmo modo que uma pessoa cega tem um sentido de tato ou de audição mais apurados. E se estendessemos as hierarquias sensoriais, ela construiria melhores modelos do mundo, como exporei em breve.

Será interessante comprovar se existe um limite máximo para o tamanho que pode atingir um sistema de memória inteligente e em que dimensões. Parece razoável pensar que o aparelho se sobrecarregaria muito para acabar sendo útil ou inclusive falharia quando chegasse a certo limite teórico. Talvez o cérebro humano já se encontra próximo do tamanho máximo teórico, mas não me parece provável. Os cérebros humanos se engrandeceram em data muito recente do período evolutivo e não há nada que sugira que nos achemos em um tamanho máximo estável. Independentemente de qual resulte ser a capacidade máxima de um sistema de memória inteligente, é quase seguro que o cérebro humano não a atingiu. É provável que nem sequer se aproxime.

Um modo de considerar o que poderiam fazer estes sistemas é observar os limites do rendimento humano conhecido. Sem dúvida, Einstein era inteligentíssimo, mas seu cérebro seguia sendo um cérebro. Cabe supor que sua extraordinária inteligência foi em boa medida produto das diferenças físicas entre seu cérebro e o cérebro humano típico. O que fez a Einstein tão raro foi que nossos genes não costumam produzir cérebros como o seu. No entanto, quando projetamos cérebros em silício, podemos os construir como quisermos. Poderiam ser capazes de atingir o elevado nível intelectual de Einstein, ou inclusive ultrapassá-lo. No outro extremo, as pessoas que padecem da síndrome do gênio (savant syndrome) poderiam lançar luz sobre outras possíveis dimensões da inteligência. Os indivíduos com a dita síndrome são atrasados mentais que mostram habilidades notáveis, como memórias quase fotográficas ou capacidade para realizar cálculos matemáticos difíceis em grande velocidade. Seus cérebros, ainda que não sejam típicos, continuam sendo cérebros e funcionam com o algoritmo cortical. Se um cérebro atípico pode possuir habilidades de memória espantosas, em teoria poderíamos acrescentar as ditas habilidades a nossos cérebros artificiais. Estes extremos da capacidade mental humana não só indicam o que deveria ser possível recriar, mas que representam a direção em que é provável que superemos os melhores resultados humanos.

Possibilidade de Duplicação

Cada novo cérebro orgânico deve crescer e treinar-se do zero, processo que leva décadas nos seres humanos. Cada humano tem de descobrir o essencial da coordenação dos membros e grupos musculares do corpo, do equilíbrio e do movimento, e aprender as propriedades gerais da multidão de objetos, animais e outras pessoas; os nomes das coisas e da estrutura da linguagem; e as regras da família e da sociedade. Uma vez que se domina esses elementos básicos, começam anos e anos de ensino formal. Todas as pessoas devem percorrer o mesmo conjunto de curvas de aprendizagem na vida —ainda que outros já as tenham transitado muitas outras inumeráveis vezes— para construir um modelo do mundo no córtex.

As máquinas inteligentes não precisam passar por esta longa curva de aprendizagem, já que os chips e o resto dos componentes podem se duplicar de forma infinita e os conteúdos se transferir com facilidade. Neste sentido, as máquinas inteligentes se duplicariam como o software. Uma vez que se pôs no ponto e treinado um sistema protótipo de forma satisfatória, o mesmo poderia ser copiado inúmeras vezes. Pode ser que leve anos de projeto de chips, configuração de hardware, treinamento e testes para aperfeiçoar o sistema de memória de um carro inteligente, mas uma vez que se conseguiu o produto final, o mesmo será produzido em série. Como já mencionei, poderíamos decidir se as cópias continuariam aprendendo ou não. Para algumas aplicações preferiremos que nossas máquinas inteligentes se limitem a operar de uma forma provada e

conhecida. Uma vez que um carro inteligente sabe tudo o que precisa, não iríamos querer que desenvolvesse maus hábitos ou chegasse a crer em alguma falsa analogia que lhe pareça ser certa. Além disso, esperaríamos que todos os carros de fabricação similar se comportassem de forma parecida. Mas para outras aplicações preferiremos que nossos sistemas de memória semelhantes ao cérebro sejam plenamente capazes de continuar aprendendo. Por exemplo, uma máquina inteligente projetada para descobrir provas matemáticas precisará da capacidade de aprender da experiência para aplicar percepções antigas a novos problemas, e em geral para ser flexível e de duração indefinida.

Será possível compartilhar componentes de aprendizagem do mesmo modo que compartilhamos componentes de software. Uma máquina inteligente de um projeto particular poderia ser reprogramada com um novo jogo de conexões que a levasse a ter um comportamento diferente, como se fosse carregado um novo conjunto de conexões em seu cérebro para que passasse de forma imediata de anglofalante a francofalante, ou de professor de ciências políticas a musicólogo. Poderia-se compartilhar e aproveitar o trabalho dos demais. Suponhamos que tenho desenvolvido e treinado uma máquina com um sistema de visão superior e outra pessoa tem desenvolvido e treinado uma máquina com um sentido da audição superior. Com o projeto adequado, poderíamos combinar o melhor de ambos sistemas sem ter que efetuar um novo treinamento de acima abaixo. Compartilhar a perícia deste modo é impossível para os humanos. A construção de máquinas inteligentes poderia evoluir seguindo as mesmas linhas que a indústria da informática, com grupos de pessoas dedicadas a treinar às máquinas inteligentes para que tenham conhecimentos e habilidades especializadas, e vendendo e compartilhando as configurações de memória resultantes. A reprogramação de uma máquina inteligente não será muito diferente da instalação de um novo videogame ou de um componente de software.

Sistemas Sensoriais

Os seres humanos contam com um punhado de sentidos que estão profundamente arraigados em nossos genes, nossos corpos e na estrutura subcortical dos nossos cérebros. Não podemos mudá-los. Às vezes empregamos a tecnologia para aumentá-los, como é o caso dos óculos de visão noturna, do radar e do telescópio espacial Hubble. Estes instrumentos de alta tecnologia são engenhosos dispositivos de interpretação de dados, não novos modos de percepção. Convertem informação que não podemos ver em indicações visuais ou auditivas que somos capazes de interpretar. De todos modos, é a espantosa flexibilidade do nosso cérebro que torna possível que olhemos a tela de um radar e compreendamos o que representa. Muitas espécies de animais possuem certos sentidos diferentes, como a ecolocalização dos morcegos e golfinhos, a capacidade das abelhas de ver a luz polarizada e ultravioleta, e o sentido de campo elétrico de alguns peixes.

Nossas máquinas inteligentes poderiam perceber o mundo através de qualquer sentido encontrado na Natureza, bem como de novos sentidos de projeto puramente humano. O sonar, o radar e a visão de infravermelho são exemplos evidentes do tipo de sentidos não-humanos que talvez gostaríamos que possuíssem as nossas máquinas inteligentes. Mas isso é apenas o começo.

Bem mais interessante é o modo como as máquinas inteligentes experimentam mundos de sensações alheios e genuinamente exóticos. Como temos visto, o algoritmo neocortical se ocupa sobretudo de encontrar padrões no mundo. Não tem preferência pelas origens físicas desses padrões. Sempre que as entradas ao córtex não sejam aleatórias e tenham uma certa riqueza ou estrutura estatística, um sistema inteligente formará memórias invariáveis e previsões a respeito. Não há razão para que estes padrões de entrada sejam análogos aos sentidos dos animais, ou se derivem do mundo real. Suspeito que é no campo dos sentidos exóticos que se encontram os usos revolucionários das máquinas inteligentes.

Por exemplo, poderíamos projetar um sistema sensorial que abranja o globo. Imaginemos sensores climáticos espaçados a cada uns cem quilômetros por continente. Estes sensores seriam análogos às células da retina. Em um dado momento, dois sensores climáticos adjacentes terão uma alta correlação em sua atividade, assim como duas células adjacentes de uma retina. Há grandes fenômenos climáticos, como as tormentas e as frentes, que se movem e mudam com o tempo, assim como há fenômenos visuais que também se comportam desta forma. Ao unir este desdobre sensorial a uma grande memória semelhante ao córtex, capacitaríamos o sistema para aprender a prever o clima, assim como você e eu aprendemos a reconhecer fenômenos visuais e a prever como evoluirão ao longo do tempo. O sistema veria padrões climáticos locais, padrões climáticos maiores e padrões que existem ao longo de décadas, anos ou horas. Ao colocar sensores bem perto em algumas regiões, criaríamos o equivalente a uma fovea, com o qual permitiríamos a nosso cérebro climático inteligente compreender e prever microclimas. Nosso cérebro climático entenderia e pensaria sobre sistemas climáticos globais como você e eu pensamos e entendemos objetos e pessoas. Os meteorólogos pretendem fazer algo similar na atualidade. Reúnem registros de diversos lugares e empregam supercomputadores para simular o clima e prever o futuro. Mas esta proposta —que é diferente no básico do modo como funcionaria uma máquina inteligente— se assemelha à forma de jogar xadrez de um computador —sem reflexão e sem entender—, enquanto nossa máquina climática inteligente está mais próxima de como joga xadrez um ser humano, pensando e compreendendo. A máquina climática inteligente descobriria padrões que se lhes têm escapado aos humanos. O fenômeno conhecido como *El Niño* não se descobriu até a década de 1960. Nosso cérebro climático poderia encontrar mais padrões como esse, ou aprender a prever tornados ou ventos muito melhor que os humanos. É difícil pôr grandes quantidades de dados sobre o clima de uma

forma que os humanos possam entender de imediato; em contraste, nosso cérebro climático sentiria e pensaria sobre o clima de forma direta.

Outros sistemas sensoriais distribuídos amplamente nos permitiriam construir máquinas inteligentes que compreendessem e predissessem as migrações animais, as mudanças demográficas e a propagação de doenças. Imaginemos que contamos com sensores distribuídos na rede de energia elétrica de um país. Uma máquina inteligente acoplada a estes sensores observaria o fluxo e o refluxo do consumo elétrico, do mesmo modo que você e eu vemos a atividade intensa do tráfego em uma estrada, ou o movimento das pessoas em um aeroporto. Mediante a exposição repetida, os humanos aprendem a prever estes padrões; basta perguntar a um empregado que vai e volta do trabalho de carro ou a um guarda de segurança do aeroporto. Da mesma forma, nosso monitor inteligente de redes elétricas seria capaz de prever demandas de energia ou situações perigosas propensas a levar a um blecaute melhor que um humano. Poderíamos combinar os sensores com o clima e a demografia humana com o fim de prever o descontentamento político, as fomes ou o surgimento de doenças. Assim como um diplomata superdotado, as máquinas inteligentes poderiam desempenhar um papel na redução do conflito e do sofrimento humanos. Talvez pense que as máquinas inteligentes precisariam de emoções para prever os padrões que envolvem comportamento humano, mas não creio nisso. Não nascemos com um jogo de cultura, um jogo de valores e um jogo de religião; tudo isso aprendemos. E do mesmo modo que posso aprender a entender as motivações de pessoas com valores diferentes dos meus, as máquinas inteligentes podem compreender as motivações e emoções humanas mesmo que elas não as tenham.

Poderíamos criar sentidos que se ocupem de entidades diminutas. Teoricamente é possível contar com sensores que possam representar padrões em células ou moléculas grandes. Por exemplo, um importante desafio atual é compreender como é possível prever a forma da molécula de uma proteína a partir da sequência de aminoácidos que compreende a dita proteína. Ser capazes de prever como se dobram e atuam as proteínas levaria ao desenvolvimento de medicinas e curas para muitas doenças. Os engenheiros e cientistas têm criado modelos de proteínas visuais tridimensionais em um esforço para prever como se comportam estas moléculas complexas, mas a tarefa tem sido muito difícil. No entanto, uma máquina superinteligente com um conjunto de sensores afinados de forma específica para esta missão talvez acharia a resposta. Se isto parece exagerado, recordemos que não nos surpreenderia se os humanos pudessem resolver o problema. Nossa incapacidade para desvendar isto talvez tenha que ver sobretudo com a falta de adequação entre os sentidos humanos e os fenômenos físicos que desejamos entender. As máquinas inteligentes podem ter sentidos criados por encomenda e uma memória maior que a humana, o que lhes permitiria resolver problemas impossíveis para nós.

Com os sentidos adequados e uma leve reestruturação da memória cortical, nossas máquinas inteligentes talvez vivam e pensem em mundos virtuais usados na matemática e na física. Por exemplo, muitos empreendimentos matemáticos e científicos requerem o entendimento do comportamento dos objetos em mundos que têm mais de três dimensões. Os teóricos das cordas, que estudam a mesma natureza do espaço, pensam em um Universo com dez ou mais dimensões. Aos seres humanos custa-lhes muito pensar sobre problemas matemáticos em quatro ou mais dimensões. Talvez uma máquina inteligente com o projeto adequado poderia compreender espaços de dimensões elevadas e, portanto, ser apta para prever como se comportam.

Por último, poderíamos unir um grupo de sistemas inteligentes em uma grande hierarquia, do mesmo modo que nosso córtex reúne a audição, o tato e a visão ascendendo na hierarquia cortical. O dito sistema aprenderia de forma automática a modelar e a prever os padrões de pensamento em populações de máquinas inteligentes. Com meios de comunicação distribuídos como a Internet, as máquinas inteligentes individuais poderiam se repartir por todo o globo. As hierarquias maiores aprendem padrões mais profundos e veem analogias mais complexas.

O objetivo destas reflexões é ilustrar que há muitos aspectos nos quais as máquinas semelhantes ao cérebro poderiam ultrapassar nossas capacidades de forma espetacular. Seriam capazes de pensar e aprender um milhão de vezes mais rápido que nós, recordar vastas quantidades de informação detalhada ou ver padrões incrivelmente abstratos. Poderiam ter sentidos mais sensíveis que os nossos, sentidos que estivessem distribuídos, ou sentidos para fenômenos muito pequenos. Poderiam pensar em três, quatro ou mais dimensões. Nenhuma destas interessantes possibilidades depende de que as máquinas inteligentes imitem aos humanos ou atuem como eles, e não envolvem uma robótica complexa.

Agora podemos ver plenamente como o teste de Turing, ao tornar equivalentes inteligência e comportamento humanos, limitou nossa visão do que é possível. Entendendo primeiro o que é a inteligência, podemos construir máquinas inteligentes que são bem mais valiosas do que a simples cópia do comportamento humano. Nossas máquinas inteligentes serão ferramentas espantosas e estenderão de forma espetacular o nosso conhecimento do Universo.

* * *

Quanto falta para que algo disto se torne realidade? Estaremos construindo máquinas inteligentes em cinquenta, vinte ou cinco anos? Há um ditado no mundo da alta tecnologia que diz que a mudança demora mais do que se espera em curto prazo, mas ocorre mais rápido do que se espera em longo prazo. Vi isto muitas vezes. Alguém se levanta em uma conferência, anuncia uma nova tecnologia e declara que estará em todos os lares em quatro anos. Acaba que o orador se

equivoca. Quatro anos convertem-se em oito, e as pessoas começam a pensar que nunca acontecerá. Justo então, quando parece que toda a ideia era um beco sem saída, começa a decolar e se converte em uma grande sensação. É provável que ocorra algo similar no negócio das máquinas inteligentes. O avanço parecerá lento no princípio, mas depois decolará com rapidez.

Nas conferências de neurociência, gosto de passear pela sala e pedir à cada participante que expresse sua opinião sobre quanto tempo passará para que tenhamos uma teoria sobre o córtex que funcione. Alguns —menos de 5 por cento— respondem “nunca” ou “já temos uma” (respostas surpreendentes, já que ganham a vida com isso). Outros 5 por cento afirma que cinco ou dez anos. A metade restante diz de dez a quinze anos, ou “na decorrência da minha vida”. Os que sobram pensam que cinquenta ou cem anos ou “nem na decorrência da minha vida”. Me coloco com os otimistas. Temos vivido em um período “lento” durante décadas, por isso a muitas pessoas lhes parece que o avanço na neurociência teórica e nas máquinas inteligentes se estancou. Se julgarmos o avanço realizado nos últimos trinta anos, será natural supor que não estamos nada próximos de uma resposta. Mas acho que encontramos-nos no momento decisivo e o assunto está a ponto de decolar.

É possível acelerar o futuro, aproximar mais o momento decisivo ao presente. Uma das metas deste livro é convencer-lhe de que, com o modelo teórico correto, podemos realizar um rápido progresso no entendimento do córtex, que com o modelo de memória-predição como guia podemos decifrar os detalhes do funcionamento e modo de pensar do cérebro. Este é o conhecimento de que precisamos para construir máquinas inteligentes. Se trabalharmos com o modelo correto, o progresso poderá avançar com rapidez.

Portanto, ainda que custa-me predizer quando se tornará realidade a era das máquinas inteligentes, acho que se hoje se dedicar gente suficiente para resolver este problema, poderemos ser capazes de criar protótipos e simulações corticais úteis em alguns anos. Antes de uma década espero que as máquinas inteligentes sejam uma das áreas mais apaixonantes da tecnologia e da ciência. Resisto-me a ser mais específico, porque sei o fácil que é subestimar o tempo que se precisa para que algo importante aconteça. Portanto, por que me mostro tão otimista a respeito do rápido avanço no entendimento do cérebro e na construção de máquinas inteligentes? Minha confiança provém em boa parte do grande tempo que já dediquei a trabalhar sobre o problema da inteligência. Quando me apaixonei pela primeira vez pelos cérebros em 1979, me pareceu que resolver o quebra-cabeça da inteligência era algo que poderia conseguir na decorrência da minha vida. Com o passar dos anos, observei com cuidado o declive da inteligência artificial, a ascensão e a queda das redes neurais e a Era do cérebro na década de 1990. Tenho visto como têm evoluído as atitudes para com a biologia teórica, e a neurociência teórica em particular. Tenho visto como têm chegado à linguagem da

neurociência as ideias de predição, representação hierárquica e tempo. Tenho visto como avançou meu entendimento e o dos meus colegas. Há dezoito anos que me interessei pelo papel da predição, e desde então de certo modo tenho vindo o provando. Como tenho estado submergido nos campos da neurociência e da informática durante mais de duas décadas, talvez meu cérebro tenha construído um modelo de alto nível sobre como ocorre a mudança tecnológica e científica, e esse modelo prediz um progresso rápido. Agora é o momento decisivo.

Epílogo

O astrônomo Carl Sagan costumava afirmar que o entendimento de algo não diminui sua admiração e mistério. Muita gente teme que o entendimento científico exige uma troca com a admiração, como se o saber retirasse o sabor e a cor da vida. Mas Sagan estava certo. A verdade é que com o entendimento nos sentimos mais cômodos em relação ao nosso papel no Universo, e ao mesmo tempo este se torna ainda mais colorido e misterioso. Ser uma mancha diminuta em um Cosmos infinito, vivo, inteligente e criativo é bem mais maravilhoso do que viver em uma Terra plana e limitada no centro de um Universo pequeno. Compreender como funcionam nossos cérebros não diminui a admiração e o mistério do Universo, das nossas vidas ou do nosso futuro. Nosso assombro não fará mais do que aumentar quando aplicarmos este conhecimento para entender a nós mesmos, construir máquinas inteligentes e assim adquirir mais conhecimento.

Ao aceitar o desafio, lembro-me do físico Erwin Schrödinger, que em 1944 escreveu um fino volume titulado *O que é a vida?* no qual convidava os jovens cientistas a ter ciência de que o funcionamento de um organismo requer leis físicas precisas e que a herança, um código escrito de algum modo pelos cromossomos, devia ser decifrável. Antes que James Watson e Francis Crick descobrissem o código genético em 1953, Schrödinger descreveu os quebra-cabeças da mutação e da entropia, e assinalou que a dita organização se mantém extraíndo ordens do meio ambiente. Muitos dos biólogos de maior sucesso do último século leram *O que é a vida?* enquanto estavam na escola ou na universidade e afirmaram que ele mudou o curso das suas vidas.

Com este livro espero incitar os jovens engenheiros e cientistas a estudar o córtex, adotar o modelo de memória-predição e construir máquinas inteligentes. Em seu ponto culminante, a inteligência artificial foi um grande movimento. Tinha revistas, programas de doutorado, livros, planos comerciais e empresários. Da mesma forma, as redes neurais suscitaram uma enorme expectativa quando a disciplina floresceu na década de 1980. Mas os modelos científicos subjacentes na inteligência artificial e nas redes neurais não eram os adequados para construir máquinas inteligentes.

Sugiro que agora contamos com um novo caminho a seguir mais promissor. Se você está na escola ou na universidade e este livro te motiva a se dedicar a esta tecnologia —construir as primeiras máquinas realmente inteligentes, e ajudar a pôr em marcha uma indústria—, te animo para que o faça. Consigamos que aconteça. Um dos truques do sucesso empresarial é que se deve saltar de cabeça primeiramente a um novo campo antes de estar totalmente seguro de que se vai conseguir o triunfo. A oportunidade é importante. Se saltar muito cedo, você entrará na briga; se esperar até que a incerteza desapareça, será muito tarde. Creio firmemente que agora é o momento de começar a projetar e construir sistemas de memória semelhantes ao córtex. Este campo será importantíssimo

científica e comercialmente. A Intel e a Microsoft da nova indústria das memórias hierárquicas se porão em marcha antes de dez anos. Realizar um empreendimento nesta escala pode parecer arriscado desde o ponto de vista financeiro ou exigente desde o intelectual, mas sempre vale a pena tentar. Espero que você se una a mim e a muitos outros que aceitam o desafio de criar uma das maiores tecnologias que o mundo jamais tenha visto.

Apêndice: Predições Verificáveis

Toda teoria deve conduzir a predições verificáveis, já que o único modo seguro de determinar a validade de uma nova ideia é a prova experimental. Por sorte, o modelo de memória-predição baseia-se na biologia e leva a várias predições específicas e novas que podem ser demonstradas. Neste apêndice enumero predições capazes de refutar ou apoiar as propostas abordadas no livro. Este material é um pouco mais avançado que o apresentado no capítulo 6 e não requer de forma alguma entender o resto do livro. Várias das predições só podem ser realizadas em animais ou em humanos acordados, porque as provas envolvem expectativa e predição do resultado de um estímulo, o que costuma requerer o estado de alerta. As predições não estão classificadas por ordem de importância.

Predição 1

Devemos encontrar células em todas as áreas do córtex, incluindo a sensorial primária, que mostrem um aumento de atividade como antecipação a um acontecimento sensorial e não como reação ao dito acontecimento sensorial.

Por exemplo, Tony Zador, do Cold Spring Harbor Laboratories, descobriu células no córtex auditivo primário dos ratos que se estimulam justo quando o rato espera escutar um som, inclusive quando não há (correspondência particular). Esta deve ser uma propriedade geral do córtex. Temos que encontrar uma atividade antecipatória similar no córtex visual e somatosensorial. As células que se ativam como antecipação de uma entrada sensorial são a definição da previsão, premissa básica do modelo de memória-predição.

Predição 2

Quanto mais específica no espaço for uma predição, mais próximos do córtex sensorial primário devemos encontrar células que se ativem adiantando a um acontecimento.

Se um macaco fosse treinado com sequências de padrões visuais de tal modo que fosse capaz de adiantar um padrão visual particular em um momento preciso, deveríamos encontrar células que mostrariam um aumento de atividade justo quando se esperava o padrão antecipado (reformulação da predição 1). Se o macaco aprendesse a ver um rosto mas não soubesse com exatidão que rosto apareceria ou como apareceria, deveríamos esperar encontrar células antecipatórias nas áreas de reconhecimento de rostos, mas não nas áreas visuais inferiores. No entanto, se o macaco se concentra em um objetivo e aprende a esperar um padrão particular em uma localização precisa do seu campo visual, deveríamos encontrar células antecipatórias em V1 ou perto. A atividade que representa a predição flui para abaixo da hierarquia cortical o mais longe que pode segundo a especificidade da predição. Às vezes consegue percorrer todo o

caminho até as áreas sensoriais primárias e em outras se detém em regiões superiores. Devem existir resultados similares em outras modalidades sensoriais.

Predição 3

As células que mostram um aumento de atividade como progresso a uma entrada sensorial deveriam se situar de modo preferencial nas camadas corticais 2,3 e 6, e a predição deveria deter seu descenso pela hierarquia nas camadas 2 e 3.

As predições que se deslocam para abaixo da hierarquia cortical o fazem através das células das camadas 2 e 3, que depois se projetam à camada 6. Estas células da camada 6 projetam-se extensamente pela camada 1 da região abaixo da hierarquia, ativando outro conjunto de células das camadas 2 e 3, e assim sucessivamente. Portanto, é nas células destas camadas (2, 3 e 6) onde devemos encontrar atividade antecipatória. Recordemos que as células ativas das camadas 2 e 3 representam um conjunto de possíveis colunas ativas; são as possíveis predições. As células ativas da camada 6 representam um número menor de colunas; são as predições específicas de uma região do córtex. Quando a predição desce pela hierarquia, a atividade acabará se detendo nas camadas 2 e 3. Por exemplo, digamos que um rato aprendeu a antecipar um de dois tons auditivos diferentes. Baseando em uma pista externa, o rato sabe quando escutará um desses dois tons, mas não é capaz de predizer qual. Neste cenário, devemos esperar ver atividade antecipatória nas camadas 2 ou 3 nas colunas que representem ambos os tons. Não tem que ter atividade na camada 6 da mesma região, porque o animal não pode predizer o tom exato que escutará. Se em outra prova o animal é capaz de predizer o tom exato que escutará, devemos ver atividade na camada 6 nas colunas que respondem a esse tom específico.

Não podemos descartar por completo a possibilidade de encontrar células antecipatórias nas camadas 4 e 5. Por exemplo, é provável que se tenha várias classes de células nestas camadas com função desconhecida. Portanto, esta predição é relativamente débil, mas continuo achando que vale a pena mencioná-la.

Predição 4

Uma classe de células das camadas 2 e 3 deveria receber de forma preferencial entradas das células da camada 6 em regiões corticais superiores.

Parte do modelo de memória-predição consiste em que as sequências aprendidas de padrões que aparecem juntas desenvolvam uma representação invariável temporariamente constante, o que chamo de "nome". Proponho que este nome é um conjunto de células das camadas 2 ou 3 que cruza uma região do córtex em diferentes colunas. O conjunto de células permanece ativo enquanto os acontecimentos que fazem parte da sequência acontecem (por exemplo, um

conjunto de células que permanece ativo enquanto se escuta uma nota de uma melodia). Este conjunto de células que representam o nome da sequência é ativado mediante realimentação das células da camada 6 das regiões superiores do córtex. Sugiro que estas células de nome são da camada 2 devido a sua proximidade com a camada 1. Mas poderia ser qualquer classe de célula das camadas 2 e 3 que tenham dendritos na camada 1. Para que funcione o sistema de nomes, os dendritos apicais dessas mesmas células de nome têm que formar sinapses preferenciais com os axônios da camada 1 que se originam na camada 6 das regiões superiores. Devem evitar formar sinapses com axônios da camada 1 que se originam no tálamo. Portanto, a teoria sugere que temos de encontrar uma classe de células dentro das camadas 2 e 3 com dendritos apicais na camada 1 que mostrem uma preferência manifesta por formar sinapses com os axônios das células da camada 6 da região superior. Outras células com sinapses na camada 1 não devem apresentar esta preferência. Até onde sei, esta é uma predição sólida e inovadora.

Uma predição e conclusão é que devemos encontrar outra classe de células nas camadas 2 ou 3 cujos dendritos apicais formem sinapses preferenciais com os axônios que se originam em regiões não específicas do tálamo. Estas células predizem os próximos elementos de uma sequência.

Predição 5

Um conjunto de células de "nome" descrito na predição 4 deveria permanecer ativo durante as sequências aprendidas.

Um conjunto de células que permanece ativo durante uma sequência aprendida é a definição de um "nome" para uma sequência predizível. Portanto, deveríamos encontrar conjuntos de células que permaneçam ativas inclusive quando a atividade das células no resto de uma coluna (células das camadas 4, 5 e 6) esteja mudando. Infelizmente, não podemos saber com precisão como será a atividade das células de nome. Por exemplo, a atividade constante de um padrão de nome poderia ser algo tão simples como um único impulso elétrico que cruzasse de forma conjunta todas as células de nome. Portanto, este grupo de células ativas poderia ser difícil de detectar.

Predição 6

Outra classe de células das camadas 2 e 3 (diferente das células de nome às que temos feito referência nas predições 4 e 5) deveriam se ativar em resposta a uma entrada imprevista, mas deveriam permanecer inativas em resposta a uma entrada adiantada.

A ideia que se encontra depois desta predição é que os acontecimentos não antecipados devem ascender pela hierarquia cortical, mas quando um

acontecimento é antecipado não queremos o passar para acima da hierarquia por que se predisse localmente. Portanto, deve existir uma classe de células nas camadas 2 e 3, diferente da classe de nome descrita nas predições 4 e 5, que mostra atividade quando ocorre um fato imprevisto, mas não quando o fato se tinha previsto. Os axônios destas células devem projetar-se à regiões superiores do córtex. Proponho um mecanismo para conseguir esta mudança de atividade. A dita célula poderia inibir-se mediante um interneurônio ativado por uma célula de nome, mas neste ponto não há modo de estabelecer uma predição sólida do mecanismo. Tudo o que cabe afirmar é que algumas células devem mostrar esta atividade diferencial. Esta é outra predição consistente e, até onde sei, inovadora.

Predição 7

Em conexão com a predição 6, os acontecimentos não antecipados deveriam se propagar ascendendo pela hierarquia. Quanto mais inovador parecer o acontecimento, mais acima deverá fluir a entrada não prevista. Os acontecimentos completamente novos teriam que chegar até o hipocampo.

Os padrões bem aprendidos se predizem mais abaixo na hierarquia, do contrário, quanto mais inovadora for uma entrada, até mais acima da hierarquia deverá se propagar. Tem que se projetar um experimento para captar esta diferença. Por exemplo, um ser humano poderia escutar uma melodia desconhecida porém simples. Se o sujeito escutar uma nota que, ainda que inesperada, é coerente com o estilo de música, a dita nota inesperada deverá causar mudanças de atividade no córtex auditivo, ascendendo até certo nível da hierarquia cortical. No entanto, se em vez de uma nota coerente com o estilo da música o sujeito escutar um som sem sentido algum, como um estouro, esperaríamos mudanças de atividade deste som para viajar mais acima na hierarquia cortical. Os resultados deveriam mudar se o sujeito estivesse esperando escutasse o estouro, mas ao invés disso escutou a nota. Se poderia provar esta predição com FMRI (imagem de ressonância magnética funcional) em sujeitos humanos.

Predição 8

O entendimento repentino deve dar como resultado uma cascata precisa de atividade preditiva que flui para abaixo da hierarquia cortical.

O momento de percepção no qual um padrão sensorial desconcertante acaba sendo entendido —como o reconhecimento da silhueta do cão dálmata da figura 11— começa quando uma região do córtex tenta uma nova correspondência da sua entrada com a memória. Se a correspondência se encaixa na região local, as predições passam para abaixo da hierarquia cortical em rápida sucessão a todas as regiões inferiores. Se trata-se de uma interpretação correta do estímulo, cada região da hierarquia estabelecerá uma predição correta em rápida sucessão. O mesmo efeito deve ocorrer enquanto se vê uma imagem com duas interpretações,

como uma silhueta em um vaso que pode parecer dois rostos ou um cubo de Necker (uma imagem de um cubo que aparece de forma alternativa em duas orientações diferentes). A cada vez que a percepção da dita imagem muda, devemos ver uma propagação de novas predições fluir para abaixo da hierarquia. Nos níveis inferiores, digamos V1, uma coluna que representa um segmento de linha da imagem deverá permanecer ativo em qualquer percepção da imagem (supondo que os olhos não tenham se movido). No entanto, poderíamos ver algumas células dessa coluna trocar estados ativos. Isto é, a mesma característica de baixo nível existe em cada imagem, mas é provável que se ativem diferentes células dentro de uma coluna segundo as interpretações diferentes. O ponto principal é que temos que ver descer a propagação das predições pela hierarquia cortical quando muda uma percepção em um nível superior.

Deve ocorrer uma propagação da predição similar com cada sacada visual sobre um determinado objeto.

Predição 9

O modelo de memória-predição requer que os neurônios piramidais possam detectar coincidências precisas de entrada sináptica em dendritos finos.

Durante muitos anos pensou-se que os neurônios poderiam ser integradores que sintetizavam as entradas de todas as suas sinapses para determinar se o neurônio deveria transmitir um impulso nervoso. Na neurociência atual há muita incerteza a respeito do comportamento dos neurônios. Algumas pessoas continuam mantendo a ideia de que são simples integradores, e muitos modelos de redes neurais se constroem com neurônios que funcionam desse modo. Mesmo assim, existem muitos modelos que assumem que um neurônio comporta-se como se cada seção dendrítica operasse de forma independente. O modelo de memória-predição requer que os neurônios sejam capazes de detectar a coincidência de algumas sinapses ativas em um lapso de tempo reduzido. O modelo poderia funcionar inclusive com uma única sinapse potencializada o suficiente para provocar com que uma célula se estimulasse, mas existiria mais probabilidades se existissem duas ou mais sinapses ativas próximas em um dendrito fino. Deste modo, um neurônio com centenas de sinapses poderia aprender a estimular muitos padrões de entrada diferentes precisos e separados. Não se trata de uma ideia nova, e há provas que a apoiam. No entanto, isto envolve um afastamento radical do modelo regular empregado durante muitos anos. Se fosse demonstrado que os neurônios não se estimulam com padrões de entrada precisos e dispersos, seria difícil manter intacta a teoria da memória-predição. As sinapses nos dendritos grossos do corpo celular ou próximas a este não precisam funcionar desta forma; só as múltiplas sinapses dos dendritos finos.

Predição 10

A extensão de árvores axonais e dendríticas deveria ampliar-se à medida que ascendemos na hierarquia cortical e afastamo-nos do córtex sensorial primário.

Tenho sustentado que V1 e outras regiões sensoriais primárias não são grandes, mas que na realidade estão compostas por muitas regiões pequenas. Estas sub-regiões encontram-se provavelmente entre as mais pequenas do córtex. Esta interpretação de V1 e outras regiões sensoriais ajuda a compreender como funciona a hierarquia, ainda que isto não contradiga muito do que cremos sobre as ditas áreas sensoriais primárias. Uma manifestação física de que existem regiões pequenas é que a expansão das árvores axonais e dendríticas deve estreitar-se no córtex sensorial primário e alargar-se à medida que se ascende de V1 a V2 e V4, por exemplo. Basicamente, esta ampliação reflete o tamanho de uma região. Já existem algumas provas experimentais que apoiam esta ideia, de modo que não é uma predição inovadora. No entanto, devemos ver esta ampliação das árvores no córtex visual, somatosensorial e motor, e é provável que também no auditivo.

Predição 11

As representações descem pela hierarquia com o treinamento.

Tenho sustentado que mediante o treinamento repetido o córtex aprenderia sequências em regiões inferiores da hierarquia cortical, o qual se deduz de forma natural do modo como a memória das sequências de padrões muda o padrão de entrada passado às próximas regiões superiores do córtex. Este processo tem um par de consequências. Uma é que devemos encontrar células que respondam a um estímulo complexo mais abaixo do córtex depois de um treinamento extenso e mais acima depois de um treinamento moderado. Em um ser humano, por exemplo, esperaríamos encontrar células que respondam às letras impressas em uma região como IT depois de um treinamento para reconhecer letras particulares. Mas após aprender a ler palavras inteiras, eu esperaria encontrar células que respondessem a letras em partes diferentes de V4, além de IT. Resultados similares devem se obter com outras espécies, regiões e estímulos. Outra consequência deste processo de aprendizagem é que se deve mudar o lugar onde se detectam lembranças e erros. Isto é, as sensações de padrões bem aprendidos devem se propagar a menos distância de ascensão pela hierarquia, o que poderia se detectar mediante técnicas de imagem. Também teríamos que ser capazes de detectar uma mudança nos tempos de reação diante de certos estímulos, porque as entradas não terão que viajar tão longe no córtex para ser reconhecidas e recordadas.

Predição 12

As representações invariáveis devem ser encontradas em todas as áreas corticais.

É bem sabido que existem células que respondem a entradas muito seletivas invariáveis aos detalhes. Observaram-se células que respondem a rostos, mãos, Bill Clinton, etc. O modelo de memória-predição prognostica que todas as regiões do córtex devem formar representações invariáveis. Estas devem refletir todas as modalidades sensoriais sob uma região do córtex. Por exemplo, se eu tivesse uma célula Bill Clinton no córtex visual, ela se estimularia a cada vez que eu visse a imagem de Bill Clinton. Se eu tivesse uma célula Bill Clinton no córtex auditivo, ela se estimularia cada vez que eu escutasse o nome "Bill Clinton". Depois esperaria encontrar células em áreas de associação que recebessem uma entrada tanto visual como auditiva que respondesse à visão ou ao nome pronunciado de Bill Clinton. Deveríamos encontrar representações invariáveis em todas as modalidades sensoriais e inclusive no córtex motor. Neste, as células representariam sequências motoras complexas. Quanto mais acima na hierarquia motora, mais complexas e invariáveis deveriam ser as representações. (Estudos recentes parecem ter descoberto células que ativam complexos movimentos da mão à boca nos macacos.) Não se trata de predições novas. A maioria dos pesquisadores creem na ideia geral de que as representações invariáveis se formam em muitas localizações por todo o córtex. No entanto, ainda que analisei isto como um fato, não se demonstrou em todas partes. O modelo de memória-predição presente que veremos as ditas células em cada uma das regiões do córtex.

* * *

As predições anteriores são algumas das formas de provar o modelo deste livro. Estou seguro de que há outras. No entanto, não é possível demonstrar se uma teoria está correta; só se está errada. Portanto, ainda que se demonstrasse que todas as predições que acabo de enumerar são certas, não constituiriam uma prova de que a hipótese da memória-predição está correta, mas seria uma prova sólida no apoio da teoria. O contrário também é verdade. Se algumas das predições precedentes acabarem não sendo verdadeiras, não invalidariam necessariamente a tese como um todo. Para algumas das predições há modos alternativos de atingir o comportamento necessário. Por exemplo, há outros modos de se criar sequências de nomes. Este apêndice só pretende mostrar que o modelo conduz a várias predições e, portanto, pode ser provado. Projetar experimentos é uma tarefa exigente e precisaria bem mais de análise do que pareça adequado para este livro. Também seria excelente encontrar modos de provar esta teoria com técnicas de imagem como FMRI. Existem muitos grandes laboratórios especializados e os ditos experimentos podem realizar-se com relativa rapidez comparado com o registro direto das células.

Bibliografia

A maioria dos livros científicos e artigos de revistas contam com extensas bibliografias que servem tanto para catalogar as contribuições dos outros como para ajudar ao leitor. Como este livro pretende chegar a um público variado no qual se incluem pessoas sem um conhecimento anterior sobre neurociência, temos evitado escrever em estilo acadêmico. Da mesma forma, esta bibliografia foi formulada antes de mais nada para que sirva de ajuda ao leitor não-expert que deseja aprender mais. Não enumero toda pesquisa importante publicada, nem tento citar a todas as pessoas que têm feito descobertas transcendentais neste campo, mas me limito a apresentar uma seleção de obras que acho que seriam bons materiais para que um leitor interessado pudesse aprender mais sobre os cérebros. Também incluo alguns textos que me acabaram sendo úteis, mas que em sua maioria são dirigidos à especialistas. Você pode encontrar análises mais profundas sobre muitos destes temas na Internet.

Infelizmente, você só encontrará algumas referências a teorias gerais sobre o cérebro, porque, como tenho assinalado no prólogo, não se escreveu muito sobre este tema, e muito menos sobre as propostas específicas deste livro.

História da Inteligência Artificial e das Redes Neurais

Arbib,

Michael A., ed., *The Handbook of Brain Theory and Neural Networks*, 2a. ed. (Cambridge, Mass.: MIT Press, 2003).

Existem muitos livros sobre as redes neurais, mas a maioria acaba não sendo muito útil para compreender o funcionamento do cérebro. Esta é uma recompilação de artigos curtos sobre vários dos diferentes temas relacionados com as redes neurais. Sua leitura proporciona uma visão geral do campo.

Baumgartner,

Peter, e Sabine Payr, eds., *Speaking Minds: Interview with Twenty Eminent Cognitive Scientists* (Princeton, N. J.: Princeton University Press, 1995).

Este livro contém entrevistas interessantes com muitos dos pensadores mais importantes da inteligência artificial, das redes neurais e da ciência cognitiva. É uma sinopse simples e agradável da história recente e do espírito do pensamento sobre a inteligência.

Dreyfus,

Hubert L., *What Computers Still Can't Do: A Critique of Artificial Reason* (Cambridge, Mass.: MIT Press, 1992).

Uma dura crítica à inteligência artificial originalmente publicada como *What Computers Can't Do* e reeditada anos depois com o título revisado. Trata-se de uma história em profundidade da inteligência artificial escrita por um dos seus mais ferrenhos críticos.

Hebb,

D. O., *The Organization of Behavior: A Neuropsychological Theory* (Nova York: Wiley, 1949).

Escrito em 1949, é o texto clássico no qual Hebb propôs a regra de aprendizagem "hebbiana" hoje aceita. Apesar de ser antigo, continua contendo algumas ideias interessantes e acaba sendo importante sobretudo desde a perspectiva histórica.

Kohonen,

Teuvo, *Self-organization and Associative Memories* (Nova York: Springer Verlag, 1984).

Para entender como funciona o córtex e como ele armazena sequências de padrões, é interessante conhecer as memórias auto-associativas. Ainda que se tenha escrito muito a respeito, não tenho encontrado fontes impressas que apresentem um resumo legível do que considero importante. Kohonen é um pioneiro neste campo. Este livro é difícil de conseguir e não é muito simples de ler, mas aborda o básico das memórias auto-associativas, incluindo a memória de sequências.

Mcculloch,

W. S., e W. Pitts, "A Logical Calculus of the Ideas Immanent in Nervous Activity", *Bulletin of Mathematical Biophysics*, vol. 5 (1943), págs. 115-133.

Artigo clássico deste autor onde propõe que os neurônios podem se conceber em essência como portas lógicas. Analiso-o no capítulo 1.

Searle,

J. R., "Minds, Brains, and Programs", *The Behavioral and Brain Sciences*, vol. 3 (1980), págs. 417-424.

Apresenta o famoso argumento da "sala chinesa" contra a computação como modelo para a mente. Pode-se encontrar muitas descrições e análises do experimento mental de Searle na Internet.

Turing,

A. M., "Computing Machinery and Intelligence", *Mind*, vol. 5 (1950), págs. 433-460.

Apresenta o famoso "teste de Turing" para detectar a presença de inteligência. Novamente, pode-se encontrar muitas referências e análises sobre o dito teste na Internet.

Neocórtex e Neurociência Geral

Os livros abaixo são recomendados para quem deseja saber mais sobre neurobiologia e sobre o neocórtex.

Crick,

Francis H. C., "Thinking about the Brain", *Scientific American*, vol. 241 (setembro 1979), págs. 181-188. Também disponível em *The Brain: A Scientific American Book* (San Francisco: W. H. Freeman, 1979).

Este é o artigo que fez com que eu me interessasse pelos cérebros. Ainda que tenha uma antiguidade de vinte e cinco anos, continua-me sendo inspirador.

Koch,

Christof, *Quest for Consciousness: A Neurobiological Approach* (Denver, Colo.: Roberts and Co., 2004).

Todos os anos se publicam vários livros sobre o cérebro de interesse geral. Este de Christof Koch trata da consciência, mas aborda a maioria dos temas importantes sobre cérebros, neuroanatomia, neurofisiologia e consciência. Se você deseja uma introdução básica à neurobiologia e à ciência cerebral em um único livro de fácil leitura, este seria um bom livro para começar.

Mountcastle,

Vernon B., *Perceptual Neuroscience: The Cerebral Cortex* (Cambridge, Mass.: Harvard University Press, 1998).

Um grande livro dedicado completamente ao neocórtex. Está muito bem escrito, tem uma divisão clara e, ainda que seja técnico, sua leitura acaba sendo agradável. É uma das melhores introduções ao neocórtex.

Kandel,

Eric R.; James H. Schwartz e Thomas M. Jessell, eds., *Principles of Neural Science*, 4a. ed. (Nova York: McGraw-Hill, 2000).

Trata-se de uma enciclopédia de um único volume sobre a neurociência. Este livro enorme não é leitura de cabeceira, mas uma boa obra de referência com que contar. Proporciona introduções detalhadas de todas as partes do sistema nervoso, incluindo neurônios, órgãos sensoriais, neurotransmissores e assim por diante.

Shepherd,

Gordon M., ed., *The Synaptic Organization of the Brain*, 5.^a ed. (Nova York: Oxford University Press, 2004).

Este livro foi-me útil, ainda que preferisse as edições anteriores, escritas por um único autor. É uma fonte técnica sobre todas as partes do cérebro, em particular as sinapses. Uso-o como referência.

Koch,

Christof, e Joel L. Davis, eds., *Large-scale Neuronal Theories of the Brain* (Cambridge, Mass.: MIT Press, 1994).

Escreveu-se muito pouco sobre as teorias gerais do cérebro. Este livro é uma compilação de artigos sobre este tema, ainda que a maioria dos que aparecem não cumpra com a meta sugerida no título. Ele proporciona uma visão geral das propostas diversas que estão sendo adotadas para compreender como funciona o cérebro em geral. Como em muitos textos recentes, pode-se encontrar elementos do modelo de memória-predição ao longo do livro.

Braitenberg,

Valentino, e Almut Schüz, *Cortex: Statistics and Geometry of Neuronal Connectivity*, 2a. ed. (Nova York: Springer Verlag, 1998).

Este livro descreve as propriedades estatísticas do cérebro do rato. Sou consciente de que isto não parece muito apaixonante, mas se trata de um livro original e útil. Conta a história do córtex em números.

Artigos Específicos Sobre Neurociência

Os seguintes artigos constituem as fontes originais de alguns dos conceitos importantes que são descritos neste livro. A maioria só pode ser encontrada em uma biblioteca universitária *on-line*.

Mountcastle,

Vernon B., "An Organizing Principle for Cerebral Function: The Unit Model and the Distributed System", em Gerald M. Edelman e Vernon B. Mountcastle, eds., *The Mindful Brain* (Cambridge, Mass.: MIT Press, 1978).

Neste artigo li pela primeira vez a proposta de Mountcastle de que o neocórtex inteiro funciona segundo um princípio comum. Ele também propõe que a coluna cortical é a unidade básica do processamento. Estas ideias são a premissa e a inspiração da teoria apresentada neste livro.

Felleman,

D. J., e D. C. Van Essen, "Distributed Hierarchical Processing in the Primate Cerebral Cortex", *Cerebral Cortex*, vol. 1 (janeiro-fevereiro 1991), págs. 1-47.

Este é o artigo clássico que descreve a organização hierárquica do córtex visual. O modelo de memória-predição baseia-se na suposição de que não só o sistema visual, mas o neocórtex inteiro, possui uma estrutura hierárquica.

Sherman,

S. M., e R. W. Guillery, "The Role of the Thalamus in the Flow of Information to the Cortex", em *Philosophical Transactions of the Royal Society of London*, vol. 357, núm. 1.428 (2002), págs. 1695-1708.

Proporciona uma visão geral da organização do tálamo e propõe a hipótese de Sherman-Guillery segundo a qual o tálamo serve para regular o fluxo de

informação entre áreas corticais. No capítulo 6 desenvolvo esta ideia no epígrafe titulado "Uma rota alternativa para ascender na hierarquia".

Rao,

R. P. e D. H. Ballard, "Predictive Coding in the Visual Cortex: A Functional Interpretation of Some Extra-Classical Receptive-field Effects", *Nature Neuroscience*, vol. 2, núm. 1 (1999), págs. 78-87.

Incluo este artigo como exemplo da recente pesquisa que fala de predição e hierarquias. Ele apresenta um modelo de realimentação nas hierarquias corticais no qual os neurônios das áreas superiores tentam predizer padrões de atividade nas áreas inferiores.

Agradecimentos

Quando alguém me pergunta como eu ganho a vida, nunca sei o que responder. A verdade é que eu ganho fazendo muito pouco, mas me rodeei de gente que parece fazer muitas coisas. Minha contribuição é chamar-lhes a atenção de vez em quando e tentar redirecionar à equipe por um novo caminho quando acaba sendo necessário. Se tenho obtido algum sucesso em minha carreira devo-o em boa parte ao trabalho duro e à inteligência dos meus colegas.

Tenho tido a sorte de conhecer a muitos cientistas e quase todos me ensinaram algo; portanto, todos eles têm contribuído às ideias que aparecem neste livro. Agradeço a todos, mas só posso mencionar alguns aqui. Bruno Olshausen, que trabalha no Redwood Neuroscience Institute (RNI) e na Universidade da Califórnia, em Davis, é uma enciclopédia ambulante sobre neurociência; ele me aponta de forma constante o que eu não sei e me sugere formas de retificar minha ignorância, o que é uma das coisas mais valiosas que alguém pode fazer. Bill Softky, também do RNI, foi a primeira pessoa que me ilustrou sobre a redução do tempo na hierarquia cortical e as propriedades dos dendritos finos. Rick Granger, da Universidade da Califórnia, em Irvine, me deu a oportunidade de compreender a memória de sequências e o papel que podia desempenhar o tálamo. Bob Knight, da Universidade da Califórnia, em Berkeley, e Christof Koch, do Califórnia Institute of Technology, têm sido fundamentais para a formação do Redwood Neuroscience Institute e muitos outros assuntos científicos. Todo o pessoal do RNI me pôs a prova e me obrigou a aperfeiçoar minhas ideias; muitas das propostas deste livro foram o resultado direto de reuniões e discussões celebradas ali. Obrigado a todos.

Donna Dubinsky e Ed Colligan têm sido meus sócios empresariais durante uma dúzia de anos. Por seu bom trabalho e ajuda consegui ser empresário enquanto trabalhava em meia jornada sobre a teoria do cérebro, algo pouco habitual. Donna costumava afirmar que um dos seus objetivos era que nossa empresa tivesse sucesso para que assim eu pudesse dedicar mais tempo à teoria do cérebro. Este livro não existiria se não fosse por Donna e Ed.

Eu não poderia ter escrito Sobre a Inteligência sem uma multidão de ajudas. Jim Levine, meu agente, acreditou no texto antes mesmo que eu soubesse o que ia escrever. Não faça um livro sem um agente como Jim. Ele apresentou-me a Sandra Blakeslee, minha coautora. Eu queria que este livro parecesse legível para um público amplo, e Sandy foi essencial para conseguir isto. Sou o culpado pelas epígrafes difíceis. Matthew Blakeslee, filho de Sandy e também escritor científico, proporcionou vários exemplos usados no texto e sugeriu o termo modelo de memória-predição. Tem sido muito agradável trabalhar com todo o pessoal da Henry Holt. Gostaria de agradecer em especial a John Sterling, diretor e editor da dita editorial. Só o vi uma vez cara a cara e falamos por telefone algumas vezes, mas foi o suficiente para que isto tivesse uma enorme repercussão sobre a estrutura do livro. Ele compreendeu de imediato os temas que eu enfrentaria ao

proponer una teoría sobre a inteligência, e despois sugeriu como o texto debía ser escrito e disposto.

Quero agradecer a minhas filhas, Anne e Kate, que non se queixaron cando seu pai pasaba moitos fins de semana diante do teclado do computador. E para finalizar, agradezo a minha esposa, Janet. Estar casada comigo non é nada fácil. Quero a ela máis do que aos cérebros.

Sobre os Autores

JEFF HAWKINS é um dos engenheiros da computação e empresários do Vale do Silício que tem atingido maior sucesso e reconhecimento. Fundador da Palm Computing, Handspring, e do Redwood Neuroscience Institute, criado para fomentar as pesquisas sobre a memória e a cognição, também é membro do conselho científico do Cold Spring Harbor Laboratories. Ele vive no norte da Califórnia.

SANDRA BLAKESKEE possui mais de trinta anos escrevendo sobre ciência e medicina para o *The New York Times* e é coautora de *Phantoms in the Brain* e dos bem sucedidos livros sobre psicologia e vida conjugal de Judith Wallerstein. Ela vive em Santa Fé, Novo México.

ÿ