
LIVRO 1
MAPA E TERRITÓRIO
RACIONALIDADE
De A a Z
ELIEZER YUDKOWSKY

RACIONALIDADE DE A a Z

MAPA E TERRITÓRIO
LIVRO 1

por ELIEZER YUDKOWSKY

Tradução de Mariana Hungria

Brasil, 2024

Sumário

Prefácio

Você está com uma compilação de dois anos de postagens diárias de um blog em suas mãos. Olhando para trás, percebo que cometi muitos erros, mas isso não me incomoda. Se não conseguisse identificar as coisas que fiz de errado, significaria que não havia melhorado minha escrita e compreensão desde 2009. Quando dizemos "opa", significa que evoluímos em nossas crenças e estratégias. Portanto, se eu olhar para trás e não ver nenhuma falha, isso significaria que eu não aprendi nem mudei minha opinião sobre nada desde então.

Cometi um erro ao não escrever minhas postagens dos últimos dois anos visando auxiliar as pessoas a terem uma vida cotidiana melhor. Em vez disso, eu me propus a ajudar a solucionar problemas importantes, difíceis e de grande porte, selecionando exemplos abstratos que soavam impressionantes.

Ao olhar para trás, percebo que esse foi o segundo maior erro na minha abordagem. Ele está relacionado ao meu maior erro na escrita, que foi não entender que o grande desafio em aprender essa valiosa forma de pensamento era descobrir como colocá-la em prática, em vez de apenas conhecer a teoria. Não me dei conta de que essa parte era a prioridade e, sobre isso, apenas posso dizer "Opa" e "Duh".

Sim, em certas ocasiões as grandes questões são realmente importantes e significativas, mas isso não muda o simples fato de que para dominar habilidades, é necessário praticá-las, e isso é mais difícil praticar utilizando coisas que estão mais distantes (atualmente, o Centro para Racionalidade Aplicada — *Center for Applied Rationality* — está trabalhando para corrigir este enorme erro que cometi de uma forma mais sistemática).

Cometi um terceiro grande erro ao me concentrar demais na crença racional e muito pouco na ação racional.

O quarto maior erro que cometi foi que eu deveria ter organizado melhor o conteúdo que eu estava apresentando nas sequências. Mais especificamente, eu deveria ter criado uma wiki muito antes, e facilitado ler as postagens em sequência. Pelo menos esse erro é corrigível. Neste trabalho, Rob Bensinger reorganizou as postagens da melhor forma possível, sem tentar reescrever o material, apesar de ter reescrito um pouco dele.

Meu quinto grande erro foi tentar, do meu ponto de vista, falar abertamente sobre o que pareciam ideias estúpidas. Eu tentei evitar o "Bulverismo", uma falácia em que se começa a discussão apontando o quão estúpidas são as pessoas por acreditarem em algo. Sempre discuti o assunto em primeiro lugar e, somente depois, dizia: "E, portanto, isso é estúpido". Contudo, em 2009, eu ainda estava incerto se seria importante atrair pessoas que expressavam desprezo pela homeopatia. Eu acreditava — e ainda acredito que existe um lamentável problema em tratar ideias com cortesia, fazendo muitas pessoas acreditarem, em algum nível, que "nada de ruim vai acontecer comigo se eu disser que acredito nisso; eu não perderei status se eu disser que acredito em homeopatia". Nesse sentido, o riso de escárnio dos comediantes pode ajudar as pessoas a despertarem do sonho.

Atualmente, eu escreveria de maneira mais cortês, acredito eu. Embora a falta de cortesia tenha desempenhado um papel útil e ajudado algumas pessoas, agora entendo melhor o risco de criar comunidades nas quais a reação normal e esperada às opiniões de baixo status é o desprezo e a zombaria. Apesar do meu erro, fico feliz em dizer que meus leitores têm sido surpreendentemente bons em não usar minha retórica

como uma desculpa para intimidar ou menosprezar os outros (gostaria de destacar Scott Alexander aqui em particular, uma pessoa mais gentil do que eu e um escritor cada vez mais excepcional sobre esses temas, e talvez mereça parte do crédito por manter a cultura saudável do Less Wrong).

Poder olhar para trás e dizer que você "fracassou" implica que você tinha objetivos. Então, o que eu estava tentando realizar?

Existe uma maneira valiosa de pensar que ainda não é ensinada nas escolas atualmente. Essa maneira de pensar não é ensinada de forma sistemática, absorvida apenas por pessoas que cresceram lendo livros como *Surely You're Joking, Mr. Feynman* (Com certeza, você está brincando, Sr. Feynman!) ou que tiveram um professor especialmente bom na escola.

Mais famosamente, essa determinada maneira de pensar tem a ver com a ciência e com o método experimental. A parte da ciência em que se sai e se observa o universo, em vez de simplesmente inventar coisas. A parte em que se diz "Opa!" e se desiste de uma teoria ruim quando os experimentos não a apoiam.

Mas essa maneira específica de pensar vai além disso. É mais profunda e mais universal do que um par de óculos que você usa quando entra em um laboratório e tira quando sai. Aplica-se à vida cotidiana, embora essa parte seja mais sutil e mais difícil. Mas se você não consegue dizer "Opa!" e desistir quando algo não está funcionando, você não terá escolha senão continuar a atirar no próprio pé. Você terá que continuar recarregando a arma e puxando o gatilho. Você conhece pessoas assim. E em algum lugar, em algum ponto da sua vida em que você prefere não pensar, você é uma pessoa assim. Seria bom se houvesse uma maneira específica de pensar que nos ajudasse a parar de fazer isso.

Mesmo com a grandeza dos meus erros, os dois anos de postagens parecem ter ajudado um número surpreendente de pessoas em um grau surpreendente. Não funcionou consistentemente, mas em alguns momentos funcionou.

Na sociedade moderna, são ensinadas tão poucas habilidades de crença racional e tomada de decisão, bem como a matemática e as ciências subjacentes a elas, que ocorre que a simples leitura de um enorme *brain-dump* (despejo mental) repleto de problemas de filosofia e ciência pode ser surpreendentemente benéfica. Analisar tudo isso a partir de uma dúzia de perspectivas diferentes pode transmitir, por vezes, uma ideia do núcleo central dessas habilidades.

Pois, no fim das contas, tudo é uma coisa só. Eu discuti grandes e importantes questões distantes e negligenciei a vida cotidiana, mas as leis que as governam não são realmente diferentes. As grandes lacunas que foquei e os exemplos escolhidos estavam errados, mas no final das contas, tudo é uma coisa só. Fico orgulhoso de olhar para trás e afirmar isso, mesmo após todos os erros que cometi e as outras vezes em que precisei dizer "Opa"…

Ainda hoje, após cinco anos, isso me parece melhor do que nada.

Eliezer Yudkowsky, fevereiro de 2015

Vieses: Uma introdução

por Rob Bensinger

Não é um segredo. No entanto, por alguma razão, é raramente mencionado nas conversas e poucas pessoas perguntam o que podemos fazer a respeito. Trata-se de um padrão, invisível, escondido por trás de todos os nossos triunfos e fracassos, por trás de nossos olhos. Qual é esse padrão?

Imagine que você coloque o braço em uma urna contendo setenta bolas brancas e trinta bolas vermelhas, e retire dez bolas de forma aleatória. Suponha que, por acaso, três das dez bolas são vermelhas e você tenta adivinhar corretamente quantas esferas vermelhas estavam na urna. Ou talvez você retire quatro bolas vermelhas, ou outro número qualquer. Nesse caso, você erraria provavelmente a quantidade total de bolas vermelhas.

Este erro aleatório é o preço a ser pago pelo conhecimento incompleto e, para um erro, não é tão ruim assim. Em média, suas estimativas não serão incorretas, e quanto mais você aprender, o erro tenderá a ser menor, menor tenderá a ser o erro.

Por outro lado, suponhamos que as bolas brancas sejam mais pesadas e afundem para o fundo da urna. Nesse caso, sua amostra pode não ser representativa o suficiente e tende a seguir em uma direção específica.

Esse tipo de erro é conhecido como “viés estatístico”. Quando o método utilizado para aprender sobre o mundo é tendencioso, obter mais informações pode não ser útil. Na verdade, coletar mais dados pode piorar consistentemente uma previsão tendenciosa.

Se você valoriza o conhecimento e a pesquisa, essa perspectiva pode ser assustadora. Se queremos ter certeza de que aprender mais nos ajudará, em vez de nos deixar em uma situação pior do que antes, precisamos identificar e corrigir os vieses em nossos dados.

Na psicologia, a ideia de viés cognitivo funciona semelhantemente. O viés cognitivo é um erro sistemático na forma como pensamos, em contraste com um erro aleatório ou um erro causado apenas por nossa ignorância. Enquanto um viés estatístico distorce uma amostra, tornando-a menos representativa de uma população maior, os vieses cognitivos distorcem nossas crenças e fazem com que elas representem com menos precisão os fatos, além de afetar nossa tomada de decisão, tornando-a menos confiável para alcançar nossos objetivos.

Pode ser que você tenha um viés de otimismo e descubra que as bolas vermelhas podem ser usadas para tratar uma doença tropical rara que afeta seu irmão. Nesse cenário, você pode superestimar o número de bolas vermelhas contidas na urna porque deseja que a maioria delas seja vermelha. Nesse caso, não é a amostra que está enviesada, mas sim você. No entanto, ao falar de pessoas enviesadas, é preciso ter cuidado.

Geralmente, quando rotulamos indivíduos ou grupos como “tendenciosos” ou “enviesados”, fazemos isso para condená-los por serem injustos ou parciais. No entanto, um viés cognitivo é algo completamente diferente. Vieses cognitivos são uma parte fundamental da maneira como os seres humanos pensam, não um tipo de defeito resultante de uma má educação ou de uma personalidade ruim. [1]

Um viés cognitivo é uma maneira sistemática pela qual seus padrões de pensamento inatos podem falhar em alcançar a verdade (ou algum outro objetivo alcançável, como a felicidade). Os vieses cognitivos, assim como os vieses estatísticos, podem distorcer nossa visão da realidade, não podendo ser facilmente corrigidos apenas recolhendo mais dados, e seus efeitos podem se acumular ao longo do tempo. No entanto, quando o instrumento de medição descalibrado que você está tentando corrigir é você mesmo, eliminar o viés se torna um desafio único.

Ainda assim, este é um ponto de partida evidente. Afinal, se você não pode confiar em seu próprio cérebro, como pode confiar em qualquer outra coisa?

Seria útil atribuir um nome a este projeto de superação de vieses cognitivos e à superação de todos os tipos de erros que nossa mente possa cometer e que possam prejudicá-la.

Podemos dar qualquer nome que desejarmos a este projeto. No entanto, por enquanto, acredito que "racionalidade" seja um nome tão adequado quanto qualquer outro.

Sentimento racional

Nos filmes de Hollywood, ser "racional" é geralmente retratado como ser uma pessoa estóica, severa e hiper intelectual. Pense no personagem Spock de Jornada nas Estrelas, que "racionalmente" suprime suas emoções, "racionalmente" se recusa a confiar em intuições ou impulsos e é facilmente confundido e enganado ao enfrentar um oponente errático ou "irracional".[2]

Existe uma compreensão completamente diferente de "racionalidade" estudada por matemáticos, psicólogos e cientistas sociais. Basicamente, trata-se da ideia de fazer o melhor possível com o que se tem. Uma pessoa racional, mesmo que esteja confusa e perdida, forma as melhores crenças possíveis com base nas evidências que possui. Uma pessoa racional, mesmo em uma situação terrível, faz as melhores escolhas possíveis para melhorar suas chances de sucesso.

A racionalidade no mundo real não consiste em ignorar suas emoções e intuições. Para um ser humano, a racionalidade muitas vezes significa tornar-se mais autoconsciente sobre seus sentimentos, de modo que possa considerá-los em suas decisões.

A racionalidade pode significar saber quando não pensar demais. Em experimentos nos quais os participantes escolheram um cartaz para colocar em sua parede ou prever o resultado de um jogo de basquete, descobriu-se que aqueles que analisaram cuidadosamente as suas razões apresentavam um desempenho pior [3][4]. Existem alguns problemas sobre os quais a deliberação consciente nos serve melhor, enquanto outros são melhor resolvidos por julgamentos momentâneos. Os psicólogos que estudam teorias de processamento duplo traçam uma distinção entre os processos cerebrais do "Sistema 1" (cognição rápida, implícita, associativa, automática) e os do "Sistema" 2 (cognição lenta, explícita, intelectual, controlada) [5]. O estereótipo é que os racionalistas confiam inteiramente no Sistema 2, desconsiderando seus sentimentos e impulsos. Olhando além do estereótipo, alguém que estivesse sendo realmente racional — realmente alcançando seus objetivos, mitigando realmente os danos de seus vieses cognitivos — dependeria fortemente dos hábitos e intuições do Sistema 1, nos casos em que eles fossem confiáveis.

Infelizmente, confiar apenas no Sistema 1 parece um mau guia para decidir "quando devo confiar no Sistema 1?" Nossas intuições sem treino não nos alertam quando devemos parar de confiar nelas. O sentimento de estar enviesado ou não é o mesmo [6].

Em compensação, como observa o economista comportamental Dan Ariely, somos previsivelmente irracionais. Cometemos erros sempre da mesma forma, de maneira repetida e sistemática.

Embora não seja possível usar nossa intuição para descobrir quando estamos caindo em um viés cognitivo, talvez possamos recorrer às ciências da mente.

As diversas faces do viés

Para solucionar problemas, nossos cérebros evoluíram para utilizar heurísticas cognitivas — atalhos rudimentares que frequentemente resultam na resposta correta, mas nem sempre. Vieses cognitivos ocorrem quando as abordagens utilizadas por essas heurísticas produzem um erro relativamente consistente e bem definido.

Por exemplo, a heurística da representatividade é a nossa tendência a avaliar fenômenos com base em quão representativos eles são de várias categorias. Isso pode levar a vieses, como a falácia da conjunção. Tversky e Kahneman descobriram que os participantes de um experimento consideravam menos provável que um jogador de tênis habilidoso "perdesse o primeiro set" do que "perdesse o primeiro set, mas ganhasse o jogo" [7]. Ganhar de virada parece mais típico de um jogador habilidoso, por isso superestimamos a probabilidade dessa narrativa complicada, mas que parece razoável, em comparação com a probabilidade de um cenário estritamente mais simples.

A heurística da representatividade também pode contribuir para negligenciarmos a frequência basal (ou taxa-base), onde tomamos decisões baseadas em quão intuitivamente "normal" uma combinação de atributos é, ignorando o quão comum esses atributos são na população em geral [8]. É mais provável que Steve seja um bibliotecário tímido ou um vendedor tímido? A maioria das pessoas responde a esse tipo de pergunta pensando se "tímido" corresponde aos estereótipos dessas profissões. Elas não consideram que os vendedores são setenta e cinco vezes mais comuns que os bibliotecários nos Estados Unidos [9].

Outros exemplos de viés incluem a negligência da duração (avaliar experiências sem considerar o tempo durado), a falácia do custo afundado (sentir-se comprometido com coisas em que você investiu recursos no passado, quando deveria limitar suas perdas e seguir em frente) e o viés de confirmação (dar mais peso às evidências que confirmam o que já acreditamos) [10] [11].

No entanto, saber sobre um viés é raramente suficiente para proteger-se dele. Em um estudo sobre cegueira para vieses, participantes previram que, se soubessem que uma pintura era obra de um artista famoso, teriam mais dificuldade em avaliar objetivamente a qualidade dela. E, de fato, aqueles informados sobre o autor da pintura e solicitados a avaliar a qualidade, exibiram o viés que eles mesmos haviam previsto, em comparação com um grupo de controle. Posteriormente, no entanto, esses mesmos indivíduos afirmaram que suas avaliações das pinturas haviam sido objetivas e não afetadas pelo viés — em todos os grupos! [12] [13]

Temos uma aversão especial em pensar que nossas opiniões são imprecisas em comparação com as dos outros. Mesmo quando identificamos corretamente os vieses dos outros, temos um ponto cego especial para nossas próprias falhas de viés [14]. Não conseguimos detectar quaisquer "pensamentos enviesados" quando refletimos internamente e, assim, concluímos que somos mais objetivos do que todas as outras pessoas [15].

De fato, estudar vieses pode torná-lo mais vulnerável ao excesso de confiança e ao viés de confirmação, porque você começa a ver a influência dos vieses cognitivos em todos ao seu redor — em todos, exceto em si mesmo. E o ponto cego, ao contrário de muitos vieses, é especialmente grave entre pessoas que são especialmente inteligentes, atenciosas e de mente aberta [16] [17].

Isso é motivo para se preocupar.

No entanto, ainda é possível melhorar. É sabido que podemos reduzir a negligência da frequência basal ao pensar em probabilidades como frequências de objetos ou eventos. Podemos minimizar a negligência da duração, prestando mais atenção à duração e representando-a graficamente [18]. As pessoas variam em quão fortemente exibem diferentes vieses, então deve haver uma série de maneiras ainda desconhecidas de influenciar o quanto somos enviesados.

No entanto, se quisermos realmente melhorar, não basta apenas estudar as listas de vieses cognitivos. A abordagem para superar esses vieses em "Racionalidade: De IA A a Z" é comunicar uma compreensão sistemática de por que o raciocínio correto funciona e como o cérebro falha nesse processo. Enquanto este livro alcança esse objetivo, sua abordagem pode ser comparada à descrita por Serfas, que observa que "anos de experiência de trabalho em finanças" não afetam a suscetibilidade das pessoas ao viés do custo afundado, enquanto "o número de cursos de contabilidade frequentados" ajuda.

Consequentemente, pode ser necessário distinguir entre experiência e perícia, sendo que perícia significa "o desenvolvimento de um princípio esquemático que envolve a compreensão conceitual do problema", o que, por sua vez, permite que o tomador de decisões reconheça determinados vieses. No entanto, usar a perícia como contramedida exige mais do que simplesmente estar familiarizado com o conteúdo da situação ou ser um especialista em um domínio específico. É preciso que a pessoa entenda completamente a lógica subjacente do respectivo viés, consiga detectá-lo no ambiente específico e também tenha à mão as ferramentas adequadas para combater o viés [19].

Este livro tem o objetivo de estabelecer as bases para criar uma "expertise" em racionalidade. Isso significa obter uma compreensão profunda da estrutura de um problema muito geral: o viés humano, a autoilusão e as muitas maneiras pelas quais o pensamento sofisticado pode se autossabotar.

Uma palavra sobre este texto

"Racionalidade: De A a Z" teve origem como uma série de ensaios escritos por Eliezer Yudkowsky, publicados entre 2006 e 2009 nos blogs de economia Overcoming Bias e Less Wrong, este último derivado do primeiro. Durante o último ano, trabalhei com Yudkowsky no Machine Intelligence Research Institute (MIRI), uma organização sem fins lucrativos fundada por ele em 2000 visando estudar os requisitos teóricos para uma inteligência artificial mais inteligente do que a humana.

Ler as postagens do seu blog despertou meu interesse pelo seu trabalho. Fiquei impressionado com a sua habilidade de comunicar concisamente insights que me levaram anos de estudo em filosofia analítica para assimilar. Na tentativa de conciliar o espírito anárquico e cético da ciência com uma abordagem rigorosa e sistemática para investigação, Yudkowsky não apenas refuta, mas também tenta entender os muitos erros e becos sem saída que a má filosofia (e a falta de filosofia) podem produzir. Ao ajudar a organizar esses ensaios em um livro, minha esperança é torná-los mais acessíveis e apreciá-los na totalidade coerente.

O manual de racionalidade resultante é frequentemente pessoal e irreverente, utilizando como ponto de partida as experiências de Yudkowsky com sua mãe (uma psiquiatra) e seu pai (um físico), ambos judeus ortodoxos, assim como conversas em salas de chat e listas de discussão. Os leitores familiarizados com Yudkowsky através de "Harry Potter e os Métodos da Racionalidade", sua abordagem cientificamente orientada da série de J.K. Rowling, vão reconhecer a mesma irreverência iconoclasta e muitos dos mesmos conceitos fundamentais.

Do ponto de vista estilístico, os ensaios presentes neste livro abrangem toda a gama de um "livro didático divertido" até um "compêndio de vinhetas inteligentes" e um "manifesto rebelde". O conteúdo é igualmente variado e abrange diversos temas. "Racionalidade: De A a Z" compila centenas de posts de Yudkowsky, organizados em vinte e seis "sequências" que tratam de temas semelhantes e funcionam como capítulos do livro. Essas sequências são agrupadas em seis livros que abrangem os seguintes tópicos:

Livro 1 - Mapa e Território: aborda a natureza das crenças e o que faz algumas delas funcionarem melhor do que outras. As quatro sequências presentes explicam conceitos bayesianos de racionalidade, crença e evidência. Um tema comum nas sequências é que o que chamamos de "explicações" ou "teorias" nem sempre funcionam como mapas precisos do mundo, o que pode levar à confusão e ao misturar nossas ideias com outras ferramentas em nossa caixa mental.

Livro 2 - Como realmente mudar sua mente: aborda a importância da verdade e questiona por que muitas vezes tiramos conclusões precipitadas e nos apegamos a nossos erros. O livro explora o motivo pelo qual é difícil adquirir crenças precisas e como podemos melhorar nesse aspecto. As sete sequências presentes discutem o raciocínio motivado e o viés de confirmação, com ênfase em como nos autoenganamos sem perceber e na armadilha de usar argumentos como soldados.

Livro 3 - A Máquina no Fantasma: questiona por que não evoluímos para ser mais racionais, mesmo com as evidências disponíveis. Apesar das limitações de recursos, parece que poderíamos estar adquirindo mais conhecimento. Para entender como e por que nossas mentes executam suas funções biológicas, precisamos examinar a evolução e o funcionamento de nossos cérebros com mais precisão. As três sequências presentes esclarecem como até filósofos e cientistas podem cometer erros quando confiam em noções evolutivas ou psicológicas intuitivas em vez de técnicas. Ao situar nossas mentes num espaço maior de sistemas guiados por objetivos, podemos identificar algumas peculiaridades do raciocínio humano e entender como tais sistemas podem "perder seu propósito".

Livro 4 - Mera Realidade: aborda a natureza do mundo em que vivemos e nosso lugar nele. Com base nos exemplos de sequências anteriores sobre como modelos evolutivos e cognitivos funcionam, as seis sequências presentes exploram a natureza da mente e das leis da física. Além de aplicar e generalizar as lições aprendidas anteriormente sobre mistérios científicos e parcimônia, esses ensaios levantam novas questões sobre o papel que a ciência deve desempenhar na promoção da racionalidade individual.

Livro 5 - Mera Bondade: Este livro explora o que confere valor — moral, estética ou prudencialmente. Essas três sequências nos convidam a refletir sobre como justificar, revisar e naturalizar nossos valores e desejos. O objetivo é encontrar uma maneira de compreender nossos objetivos sem comprometer nossos esforços para alcançá-los efetivamente. Aqui, o maior desafio reside em saber quando confiar em nossos impulsos confusos e complexos sobre o que é certo e errado em cada situação, e quando os substituir por princípios simples e sem exceção.

Livro 6 - Tornando-se mais forte: este livro aborda como indivíduos e comunidades podem colocar tudo isso em prática. Essas três sequências iniciam com um relato autobiográfico dos maiores equívocos filosóficos do próprio Yudkowsky, acompanhados de conselhos sobre como ele acredita que os outros possam fazer melhor. O livro conclui com recomendações para o desenvolvimento de currículos baseados em evidências, voltados à racionalidade aplicada, bem como para a criação de grupos e instituições que apoiem estudantes interessados, educadores, pesquisadores e amigos.

As sequências são complementadas por "interlúdios", ensaios extraídos do website pessoal de Yudkowsky. Esses ensaios se conectam às sequências de várias maneiras. Por exemplo, "As Doze Virtudes da Racionalidade" resume poeticamente muitas das lições abordadas em "Racionalidade: De A a Z" e é frequentemente citado em outros ensaios.

Ao final de cada ensaio, você encontrará um asterisco. Clicando nele, você será direcionado para sua versão original no Less Wrong (onde é possível deixar comentários) ou no site de Yudkowsky. Além disso, você pode encontrar um glossário online com terminologia relacionada à "Racionalidade: de A a Z" em http://wiki.lesswrong.com/wiki/RAZ_Glossary.

Mapa e território

Este primeiro livro inicia com uma sequência sobre vieses cognitivos intitulada "Previsivelmente Errado". No entanto, o escopo do livro é mais amplo. Maus hábitos e ideias equivocadas são igualmente importantes, mesmo quando originados do conteúdo de nossas mentes e não da estrutura em si. Assim, tanto erros evoluídos como aqueles inventados serão abordados nessas sequências. O livro começa explorando, em "Crenças falsas", como as expectativas de uma pessoa podem se distanciar das crenças que ela professa.

Um relato abrangente sobre irracionalidade seria incompleto sem uma teoria sobre como a racionalidade funciona. Afinal, não bastaria uma teoria composta apenas de obviedades vagas, sem mecanismos

explicativos precisos. Por isso, a sequência intitulada “Percebendo Confusão” explora por que é vantajoso fundamentar o comportamento em expectativas “racionais” e qual é a sensação de agir dessa maneira.

A seguir, a sequência “Respostas Misteriosas” questiona se a ciência resolve esses problemas para nós. Os cientistas baseiam seus modelos em experimentos reproduzíveis, e não em especulações ou boatos. A ciência tem um histórico excelente quando comparada a anedotas, religião e praticamente tudo o mais. Entretanto, mesmo com essas características, ainda precisamos nos preocupar com crenças “falsas”, viés de confirmação, viés de retrospectiva e similares, especialmente ao lidar com uma comunidade de pessoas que buscam explicar fenômenos em vez de apenas contar histórias atraentes.

Esta sequência é seguida por “A simples verdade”, uma alegoria independente sobre a natureza do conhecimento e da crença.

Contudo, é o viés cognitivo que oferece a visão mais clara e direta sobre a natureza de nossa psicologia, manifestando-se por meio de nossas heurísticas e das limitações de nossa lógica. Por isso, é com o viés que começaremos nossa exploração.

Há uma passagem no Zhuangzi, um texto filosófico proto-taoísta, que diz: “A armadilha para peixes existe devido ao peixe; uma vez que você tenha obtido o peixe, você pode esquecer a armadilha.” [20]

Convido-o a abordar este livro com essa mentalidade. Use-o da mesma forma que usamos uma armadilha para peixes, mantendo sempre em mente o objetivo que você tem para ele. Leve consigo aquilo que possa ser útil, enquanto for relevante; descarte o que não for necessário. Que o propósito do livro o sirva bem ao longo de sua jornada.

Para ver a seção de agradecimentos, notas e bibliografia, consulte o texto original.

Agradecimentos

Sou imensamente grato a Nate Soares, Elizabeth Tarleton, Paul Crowley, Brienne Strohl, Adam Freese, Helen Toner e dezenas de voluntários que revisaram partes deste livro.

Gostaria de expressar meus mais sinceros agradecimentos a Alex Vermeer, que supervisionou a produção deste livro do início ao fim, e a Tsvi Benson-Tilsen, que revisou minuciosamente cada página para garantir sua legibilidade e consistência.

Notas

1. A ideia do viés pessoal, viés da imprensa, e assim por diante, assemelha-se ao viés estatístico por ser um erro. Outras maneiras de generalizar a ideia de "viés" focam em sua associação com a não aleatoriedade. No contexto do aprendizado de máquina, por exemplo, um viés indutivo é simplesmente o conjunto de suposições que um agente utiliza para fazer previsões com base em um conjunto de dados. Nesse caso, o agente está "enviesado" no sentido de que é direcionado em uma direção específica. No entanto, uma vez que essa direção pode representar a verdade, não é necessariamente negativo para um agente ter um viés indutivo**.** Isso pode ser valioso e necessário. Essa característica distingue claramente o viés indutivo dos outros tipos de viés.

2. Uma triste coincidência ocorreu: Leonard Nimoy, o ator que interpretou Spock, faleceu poucos dias antes do lançamento deste livro. Embora tenhamos mencionado seu personagem como um exemplo clássico da falsa 'racionalidade de Hollywood', queremos deixar claro que não temos a intenção de desrespeitar, de forma alguma, a memória de Nimoy.

3. Timothy D. Wilson et al., "Introspecting About Reasons Can Reduce Post-choice Satisfaction,"Personality and Social Psychology Bulletin 19 (1993): 331–331.

4. Jamin Brett Halberstadt and Gary M. Levine, "Effects of Reasons Analysis on the Accuracy of Predicting Basketball Games," Journal of Applied Social Psychology 29, no. 3 (1999): 517–530.

5. Keith E. Stanovich and Richard F. West, "Individual Differences in Reasoning: Implications for the Rationality Debate?," Behavioral and Brain Sciences 23, no. 5 (2000): 645–665, http://journals. cambridge. org/abstract_S0140525X00003435.

6. Timothy D. Wilson, David B. Centerbar, and Nancy Brekke, "Mental Contamination and the Debiasing Problem," in Heuristics and Biases: The Psychology of Intuitive Judgment, ed. Thomas Gilovich, Dale Griffin, and Daniel Kahneman (Cambridge University Press, 2002).

7. Amos Tversky and Daniel Kahneman, "Extensional Versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment," Psychological Review 90, no. 4 (1983): 293–315, doi:10.1037/0033-295X.90.4.293.

8. Richards J. Heuer, Psychology of Intelligence Analysis (Center for the Study of Intelligence, Central Intelligence Agency, 1999).

9. Wayne Weiten, Psychology: Themes and Variations, Briefer Version, Eighth Edition (Cengage Learning, 2010).

10. Raymond S. Nickerson, "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises," Review of General Psychology 2, no. 2 (1998): 175.

11. A negligência da probabilidade é outro viés cognitivo. Nos meses e anos seguintes aos ataques de 11 de setembro, muitas pessoas optaram por fazer longas viagens de carro em vez de voar. Embora o sequestro não fosse provável, agora parecia uma possibilidade, e essa mera possibilidade afetou profundamente as decisões. Ao confiar em um raciocínio simplista (carros e aviões são "seguros" ou "inseguros", sem meio-termo), as pessoas, na verdade, colocaram-se em um perigo muito maior. Em vez de avaliar a probabilidade de morrer em uma viagem de carro pelo país em comparação com a probabilidade de morrer em um voo pelo país — sendo as primeiras centenas de vezes mais provável — elas confiaram em seu sentimento geral de preocupação e ansiedade (a heurística do afeto). Podemos observar o mesmo padrão de comportamento em crianças que, ao ouvir os argumentos a favor e contra a segurança dos cintos de segurança, oscilam entre acreditar que cintos de segurança são completamente bons ou completamente ruins, em vez de tentar comparar o peso dos argumentos a favor e contra. [21]

Alguns outros exemplos de vieses são:

A regra do pico/final (avaliar eventos relembrados com base em seu momento mais intenso, e em como eles acabaram); [22]

Ancoragem (basear suas decisões em informações recentemente encontradas, mesmo quando irrelevantes);

Autoancoragem (usar a si próprio como um modelo de características prováveis dos outros sem considerar suficientemente os sentidos nos quais você é atípico); [23]

O viés do status quo (favorecer excessivamente o que é normal e esperado em detrimento do que é novo e diferente). [24]

12. Katherine Hansen et al., "People Claim Objectivity After Knowingly Using Biased Strategies," Personality and Social Psychology Bulletin 40, no. 6 (2014): 691–699.

13. Da mesma forma, Pronin escreve sobre a cegueira por preconceito de gênero:

Em um estudo, os participantes foram apresentados a um candidato e uma candidata a uma posição de chefe de polícia e, em seguida, foram solicitados a avaliar se "ser malandro" ou "ter educação formal" era mais importante para o trabalho. O resultado revelou que os participantes favoreceram a característica que lhes foi informada que o candidato do sexo masculino possuía (por exemplo, se lhes disseram que ele era "malandro", consideravam essa característica mais importante). Os participantes estavam completamente inconscientes desse viés de gênero; de fato, quanto mais objetivos eles acreditavam ser, maior era o viés que realmente demonstravam. [25]

Mesmo quando temos conhecimento sobre vieses, Pronin observa que ainda somos "realistas ingênuos" em relação às nossas próprias crenças. De maneira consistente, continuamos a considerar nossas crenças como representações fiéis e não distorcidas da realidade, mesmo quando isso pode não ser verdade. [26]

14. Em uma pesquisa realizada com 76 pessoas que aguardavam nos aeroportos, os indivíduos classificaram-se como sendo muito menos suscetíveis a vieses cognitivos, em média, do que uma pessoa comum no aeroporto. Especificamente, as pessoas tendem a considerar-se especialmente imparciais quando o viés é socialmente indesejável ou quando suas consequências são difíceis de serem percebidas. [27] Outros estudos também constataram que as pessoas que possuem ligações pessoais com um determinado problema acreditam que essas ligações melhoram sua compreensão e objetividade. No entanto, quando essas mesmas pessoas observam outras com as mesmas ligações, elas inferem que essas pessoas estão excessivamente envolvidas e tendenciosas.

15. Joyce Ehrlinger, Thomas Gilovich, and Lee Ross, "Peering Into the Bias Blind Spot: People's Assessments of Bias in Themselves and Others," Personality and Social Psychology Bulletin 31, no. 5 (2005): 680–692.

16. Richard F. West, Russell J. Meserve, and Keith E. Stanovich, "Cognitive Sophistication Does Not Attenuate the Bias Blind Spot," Journal of Personality and Social Psychology 103, no. 3 (2012): 506.

17. Não devem ser confundidas com aquelas pessoas que acreditam ser especialmente inteligentes, atenciosas, entre outras qualidades, devido ao viés da superioridade ilusória.

18. Michael J. Liersch and Craig R. M. McKenzie, "Duration Neglect by Numbers and Its Elimination by Graphs," Organizational Behavior and Human Decision Processes 108, no. 2 (2009): 303–314.

19. Sebastian Serfas, Cognitive Biases in the Capital Investment Context: Theoretical Considerations and Empirical Experiments on Violations of Normative Rationality (Springer, 2010).

20. Zhuangzi and Burton Watson, The Complete Works of Zhuangzi (Columbia University Press, 1968).

21. Cass R. Sunstein, “Probability Neglect: Emotions, Worst Cases, and Law,” Yale Law Journal (2002): 61–107.

22. Dan Ariely, Predictably Irrational: The Hidden Forces That Shape Our Decisions (HarperCollins, 2008).

23. Boaz Keysar and Dale J. Barr, “Self-Anchoring in Conversation: Why Language Users Do Not Do What They ‘Should,”’ in Heuristics and Biases: The Psychology of Intuitive Judgment: The Psychology of Intuitive Judgment, ed. Griffin Gilovich and Daniel Kahneman (New York: Cambridge University Press, 2002), 150–166, doi:10.2277/0521796792.

24. Scott Eidelman and Christian S. Crandall, “Bias in Favor of the Status Quo,” Social and Personality Psychology Compass 6, no. 3 (2012): 270–281.

25. Eric Luis Uhlmann and Geoffrey L. Cohen, “‘I think it, therefore it’s true’: Effects of Self-perceived Objectivity on Hiring Discrimination,” Organizational Behavior and Human Decision Processes 104, no. 2 (2007): 207–223.

26. Emily Pronin, “How We See Ourselves and How We See Others,” Science 320 (2008): 1177–1180, http://psych.princeton.edu/psychology/research/pronin/pubs/2008%20Self%20and%20Other.pdf.

27. Emily Pronin, Daniel Y. Lin, and Lee Ross, “The Bias Blind Spot: Perceptions of Bias in Self versus Others,” Personality and Social Psychology Bulletin 28, no. 3 (2002): 369–381.

A — Previsivelmente errado

1 — O que quero dizer com “Racionalidade”?

Quero dizer:

1. **Racionalidade epistêmica**: aprimorar sistematicamente a precisão das suas crenças.
2. **Racionalidade instrumental**: alcançar sistematicamente os seus valores.

Ao abrir os olhos e olhar ao redor do quarto, você localiza seu laptop em relação à mesa e identifica a estante de livros em relação à parede. No entanto, se houver algum problema com seus olhos ou com o seu cérebro, seu modelo mental pode falhar. Você pode pensar que há uma estante de livros onde, na verdade, não existe. E quando for até lá para pegar um livro, ficará desapontado.

Ter uma falsa crença é como ter um mapa do mundo que não corresponde à realidade. A racionalidade epistêmica, por outro lado, visa construir mapas precisos dessa realidade. A correspondência entre crença e realidade é o que comumente chamamos de “verdade”, e eu não vejo nenhum problema em usar esse termo.

A racionalidade instrumental, por outro lado, trata de direcionar a realidade—conduzir o futuro para onde você quer que ele vá. É a arte de escolher ações que levam a resultados melhores, segundo as suas preferências. Eu às vezes me refiro a isso como “vencer”.

Portanto, racionalidade é formar crenças verdadeiras e tomar decisões vencedoras. Buscar a “verdade” aqui não significa ignorar evidências incertas ou indiretas. Olhar ao redor do quarto e construir um mapa mental dele não é diferente, em princípio, de acreditar que a Terra tem um núcleo derretido ou que Júlio César era calvo. Essas perguntas, embora distantes no espaço e no tempo, não são menos concretas ou relevantes do que a localização de sua estante. Existem fatos sobre o estado do núcleo da Terra no ano 2015 EC[1] e sobre a aparência de César em 50 AEC[2]. Esses fatos podem impactar a sua vida de maneiras significativas, mesmo que você nunca viaje para o centro da Terra ou volte no tempo.

E “vencer” neste contexto não implica necessariamente em superar ou derrotar outros. A vida pode ser sobre colaboração e autossacrifício, ao invés de competição. “Seus valores” se referem a qualquer coisa que você considere importante, incluindo outras pessoas. Não se limita a valores egoístas ou valores não compartilhados.

Quando as pessoas dizem “X é racional!” geralmente o que elas estão enfatizando é que acreditam que X é “verdadeiro” ou “bom”. Então, por que ter uma palavra para “racional”, além das palavras “verdadeiro” e “bom”?

Um argumento análogo pode ser apresentado contra o uso da palavra “verdade”. Não há necessidade de dizer “é verdade que a neve é branca” quando você poderia apenas dizer “a neve é branca”. O que torna o conceito de verdade em algo útil é que ele nos permite falar sobre os aspectos gerais da correspondência mapa-território. “Modelos verdadeiros geralmente produzem previsões experimentais melhores do que modelos falsos” — essa é uma generalização útil, e não é uma afirmação que você possa fazer sem usar um conceito como “verdadeiro” ou “exato”.

---

1 NT.: Era Comum

2 NT.: Antes da Era Comum

Da mesma forma, “Agentes racionais tomam decisões que maximizam a expectativa probabilística de uma função de utilidade coerente” é o tipo de pensamento que depende de um conceito de racionalidade (instrumental), enquanto “Comer vegetais é uma escolha racional” provavelmente pode ser substituído por “Comer vegetais é uma escolha útil” ou “É do seu interesse comer vegetais”.

Precisamos de um conceito de ‘racional’ para identificar padrões de pensamento que levam à verdade ou geram valor. Além disso, esse conceito nos ajuda a compreender as formas sistemáticas pelas quais podemos falhar em atingir esses padrões.

Às vezes, os psicólogos experimentais descobrem raciocínios humanos que parecem muito estranhos. Por exemplo, alguém classifica a probabilidade de “Bill toca jazz” como menor do que a probabilidade de “Bill é um contador que toca jazz”. Isso parece um julgamento estranho, já que qualquer contador, em particular, que toca jazz, é, obviamente, alguém que toca jazz. Mas quando dizemos que esse julgamento está errado, a que perspectiva superior estamos apelando? Os psicólogos experimentais usam dois exames de referência: a teoria da probabilidade e a teoria da decisão.

A teoria da probabilidade fornece as leis fundamentais que regem a crença racional. A matemática por trás dela descreve, de forma igual e indistinta, três situações: (a) localizar sua estante de livros, (b) determinar a temperatura do núcleo da Terra e (c) estimar a quantidade de cabelos que Júlio César tinha. O desafio em todas essas situações é o mesmo: processar as evidências e observações para atualizar nossas crenças da maneira mais precisa possível. De forma análoga, a teoria da decisão fornece as leis fundamentais para a ação racional, e essas leis são aplicáveis independentemente dos objetivos e opções disponíveis.

Proponho representarmos “P (tal e tal)” como “a probabilidade de que tal e tal aconteça” e P (A, B) como “a probabilidade de que A e B aconteçam”. Uma lei universal da teoria da probabilidade é que P (A) ≥ P (A, B). Portanto, seria incorreto afirmar que P(Bill toca jazz) é menor do que P (Bill toca jazz e Bill é contador). Usando um jargão técnico, você diria que esse juízo de probabilidade não é bayesiano. Crenças e ações que são racionais nesse sentido matematicamente bem definido são chamadas de “Bayesianas”.

Observe que o conceito moderno de racionalidade não se limita ao raciocínio por meio de palavras. Um exemplo disso é quando abrimos os olhos, observamos o que nos cerca e criamos um modelo mental de um quarto que contém uma estante encostada na parede. O conceito moderno de racionalidade é amplo o suficiente para incluir nossos olhos e as áreas visuais do nosso cérebro como componentes que ajudam a criar esse modelo mental. Ele também engloba nossas intuições não verbais. Na matemática, não importa se usamos a mesma palavra em inglês, “racional”, para nos referirmos tanto a Spock quanto ao Bayesianismo. As equações matemáticas representam maneiras eficazes de alcançar objetivos e compreender o mundo, independentemente de essas maneiras coincidirem com nossas ideias pré-concebidas e estereótipos sobre o que é “racional”.

Isso não esgota o problema do que se entende na prática por “racionalidade”, por duas razões principais:

Primeiramente, os formalismos bayesianos em sua forma completa são computacionalmente intratáveis na maioria dos problemas reais. Ninguém pode realmente calcular e obedecer à matemática, assim como você não consegue prever o mercado de ações calculando os movimentos dos quarks.

É por isso que existe um website inteiro chamado *Less wrong* (Menos errado), em vez de uma única página que simplesmente declara os axiomas formais e encerra o assunto. Existe toda uma arte adicional para encontrar a verdade e alcançar valor dentro da nossa própria mente: precisamos aprender sobre nossas próprias falhas, superar nossos preconceitos, evitar nos autoenganar, colocar-nos em boa forma emocional para enfrentar a verdade e fazer o que precisa ser feito, e assim por diante.

Em segundo lugar, às vezes o significado da própria matemática é questionado. As regras precisas da teoria da probabilidade, são questionadas, como no exemplo dos problemas antrópicos, onde o número de observadores é incerto. Da mesma forma, as regras exatas da teoria da decisão são questionadas por problemas como o de Newcomb, nos quais outros agentes podem prever a sua decisão antes mesmo que você a tome.

Em casos como esses, é inútil tentar resolver o problema apresentando uma nova definição da palavra "racional" e dizendo: "Portanto, minha resposta preferida, por definição, é o significado da palavra 'racional'." Isso simplesmente levanta a questão de porque alguém deveria prestar atenção à sua definição. Não estou interessado na teoria da probabilidade porque é a palavra sagrada transmitida por Laplace. Estou interessado em atualizar crenças no estilo Bayesiano (com priori de Ocam) porque espero que este estilo de pensamento nos leve sistematicamente mais perto, você sabe, precisão, o mapa que reflete o território.

Existem perguntas sobre como pensar que parecem não ser totalmente respondidas nem pela teoria da probabilidade, nem pela teoria da decisão. Um exemplo é a questão de como se sentir em relação à verdade após descobri-la. Tentar definir "racionalidade" de uma maneira particular não fornece uma resposta, mas apenas pressupõe uma.

Não estou aqui para discutir o significado de uma palavra, nem mesmo se essa palavra for "racionalidade". O objetivo de vincular sequências de letras a conceitos específicos é permitir que duas pessoas se comuniquem — ajudar a transportar pensamentos de uma mente para outra. Você não pode mudar a realidade, ou provar o pensamento, manipulando quais significados acompanham quais palavras.

Portanto, se você entender a que conceito geralmente estou me referindo com esta palavra "racionalidade" e com os subtermos "racionalidade epistêmica" e "racionalidade instrumental", nós nos comunicamos: alcançamos tudo o que há para alcançar, falando sobre como definir "racionalidade". O que resta discutir não é qual significado atribuir às sílabas" ra-cio-na-li-da-de; o que resta discutir é o que é uma boa maneira de pensar.

Se você diz: "É (epistemologicamente) racional, para mim, acreditar em X, mas a verdade é Y", provavelmente está usando a palavra "racional" com um significado diferente do que tenho em mente. (Por exemplo, "racionalidade" deve ser consistente sob reflexão — "racionalmente" olhando para a evidência, e "racionalmente" considerando como sua mente processa a evidência, não deve levar a duas conclusões diferentes.)

Da mesma forma, se você se pegar dizendo: "A coisa (instrumentalmente) racional a se fazer é X, mas a coisa certa a se fazer é Y", então você quase certamente está usando algum outro significado para a palavra "racional" ou para a palavra "certa". Emprego o termo "racionalidade" normativamente, para escolher padrões de pensamento desejáveis.

Neste caso — ou em qualquer outro caso em que as pessoas discordem sobre o significado das palavras — você deve utilizar uma palavra mais específica no lugar de "racional": "A coisa benéfica a fazer é fugir, mas espero pelo menos tentar arrastar a criança para fora dos trilhos da ferrovia" ou "A teoria da decisão causal, como geralmente formulada, diz que você deve apostar duas caixas no Problema de Newcomb, mas Prefiro um milhão de dólares."

Na verdade, recomendo a leitura deste ensaio, substituindo todas as instâncias de "racional" por "tolo" e ver se isso muda as conotações do que estou dizendo. Se assim for, eu digo: lute não pela racionalidade, mas pela tolice. A palavra "racional" tem potenciais armadilhas, mas há muitos casos não limítrofes em que "racional" funciona bem para comunicar o que quero dizer. Da mesma forma, "irracional". Nesses casos, não tenho medo de usá-la.

No entanto, deve-se ter cuidado para não abusar dessa palavra. Não se recebe pontos apenas por pronunciá-la em voz alta. Se você falar demais do Caminho, não o alcançará.

Referências

1. *[Nota do editor: para uma boa introdução ao Problema de Newcomb, consulte Holt. De forma mais geral, você pode encontrar definições e explicações para muitos dos termos deste livro no site wiki. lesswrong.com/wiki/RAZ_Glossary.]*
2. *Jim Holt, "Thinking Inside the Boxes," Slate (2002), http://www.slate.com/articles/arts/egghead/2002/02/thinkinginside%5C_the%5C_boxes.single.html.*

2 — Sentimento racional

Uma crença popular sobre racionalidade é que ela se opõe a toda emoção — que toda nossa tristeza e alegria são automaticamente antilógicas por serem sentimentos. No entanto, estranhamente, não consigo encontrar nenhum teorema da teoria da probabilidade que prove que devo parecer frio e inexpressivo.

Então, a racionalidade é independente do sentimento? Não, nossas emoções surgem de nossos modelos de realidade. Se eu acreditar que meu irmão morto foi encontrado vivo, ficarei feliz; se eu acordar e perceber que foi um sonho, ficarei triste. P.C. Hodgell disse: "Aquilo que pode ser destruído pela verdade deve ser." A felicidade do meu eu sonhador foi destruída pela verdade. Minha tristeza ao acordar é racional; não há verdade que a destrua.

A racionalidade começa perguntando como o mundo é, mas se espalha para qualquer outro pensamento que dependa de como pensamos que o mundo é. Suas crenças sobre "como-o-mundo-é" podem dizer respeito a qualquer coisa que você acha que existe na realidade, qualquer coisa que exista ou não, qualquer membro da classe "coisas que podem fazer outras coisas acontecerem". Se você acredita que existe um duende em seu armário que amarra os cadarços de seus sapatos, então essa é uma crença sobre como o mundo é. Seus sapatos são reais — você pode pegá-los. Se há algo lá fora que pode alcançar e amarrar seus cadarços, isso deve ser real, também, parte da vasta rede de causas e efeitos que chamamos de "universo".

Sentir raiva do duende que amarrou seus cadarços envolve um estado de espírito que não é apenas sobre como o mundo é. Suponha que, como um budista, um paciente de lobotomia ou apenas uma pessoa muito calma, encontrar seus cadarços amarrados não o deixe com raiva. Isso não afetaria o que você espera ver no mundo — você ainda esperaria abrir seu armário e encontrar seus cadarços amarrados. Seu estado emocional não deve afetar seu melhor palpite aqui, porque o que acontece em seu armário não depende de seu estado emocional; embora possa ser necessário algum esforço para pensar com clareza.

Mas o sentimento de raiva está emaranhado com um estado de espírito sobre como o mundo é; você fica com raiva porque acha que o duende amarrou seus cadarços. O critério da racionalidade se espalha viralmente, desde a pergunta inicial se um duende amarrou ou não seus cadarços, até a raiva resultante.

Tornar-se mais racional — chegar a melhores estimativas de como o mundo é — pode diminuir ou intensificar sentimentos. Às vezes, fugimos de sentimentos fortes, negando os fatos, evitando a visão de mundo que deu origem à emoção poderosa. Nesse caso, à medida que você estuda as habilidades da racionalidade e se treina para não negar os fatos, seus sentimentos se fortalecem.

Quando eu era mais novo, eu nunca tinha certeza se era certo sentir as coisas intensamente — se era permitido, se era apropriado. Não acredito que essa confusão tenha surgido apenas da minha incompreensão juvenil da racionalidade. Notei problemas semelhantes em pessoas que nem mesmo aspiram a ser racionalistas; quando estão felizes, eles se perguntam se realmente podem estar felizes e, quando estão tristes, nunca têm certeza se devem fugir da emoção ou não. Desde os dias de Sócrates, pelo menos, e provavelmente muito antes, a maneira de parecer culto e sofisticado é nunca deixar ninguém ver que você se preocupa muito com qualquer coisa. É vergonhoso sentir, isso simplesmente não é feito na sociedade bem-educada. Você deveria ver os olhares estranhos que recebo quando as pessoas percebem o quanto me preocupo com a racionalidade. Não é um assunto incomum, eu acho, mas eles não estão acostumados a ver adultos saudáveis que visivelmente se importam com qualquer coisa.

Mas sei agora que não há nada de errado em sentir intensamente. Desde que adotei a regra de “Aquilo que pode ser destruído pela verdade, assim deve ser”, também percebi que “Aquilo que a verdade nutre deve prosperar”. Quando algo bom acontece, fico feliz, e não há confusão em minha mente se é racional eu estar feliz. Quando algo terrível acontece, não fujo da minha tristeza procurando consolações falsas e ilusórias. Com cada dia que passa, contemplo a trajetória da humanidade, desde o passado até o futuro. Reflito sobre as inúmeras vidas que se perderam ao longo dos tempos, sobre a miséria e o medo que acompanharam nossa existência. Observo as mãos trêmulas que se erguem em meio a tanto sangue derramado. Imagino o que seremos capazes de nos tornar quando as estrelas se tornarem nossas cidades. Toda a escuridão e toda a luz que fazem parte de nossa história e que jamais poderei entender completamente. Não tenho palavras para descrever essa jornada, mas a contemplo com admiração e respeito. Apesar de toda a minha filosofia, ainda sinto vergonha de confessar emoções intensas e provavelmente você não se sentirá confortável em ouvi-las. Mas sei agora que é racional sentir.

3 — Por que verdade? E...

Alguns dos comentários no blog *Overcoming Bias* (Superando o viés) abordaram a questão do porquê devemos buscar a verdade. (Felizmente, muitos não questionaram) Nossa motivação para configurar nossos pensamentos para a racionalidade, que determina se uma configuração em particular é "boa" ou "ruim", vem do motivo pelo qual queríamos encontrar a verdade em primeiro lugar.

Está escrito: "A primeira virtude é a curiosidade". A curiosidade é um motivo para buscar a verdade, e pode não ser o único, mas tem uma pureza especial e admirável. Se seu motivo for a curiosidade, você atribuirá prioridade às perguntas como as próprias perguntas agradam ao seu senso estético pessoal. Um desafio mais complicado, com maior probabilidade de falha, pode valer mais esforço do que um mais simples, apenas porque ele é mais divertido.

Como observei, as pessoas costumam pensar na racionalidade e na emoção como adversários. Como a curiosidade é uma emoção, suspeito que algumas pessoas se oporão ao tratamento da curiosidade como parte da racionalidade. De minha parte, rotulo uma emoção como não racional se ela se baseia em crenças errôneas, ou melhor, em uma conduta epistêmica produtora de erros: se o ferro se aproxima do seu rosto e você acredita que está quente, mas, na verdade, está frio, o Caminho se opõe ao seu medo. Se o ferro se aproxima do seu rosto e você acredita que ele está frio, mas, na verdade, ele está quente, o Caminho se opõe à sua calma. Inversamente, então, uma emoção evocada por crenças corretas ou pensamento epistemologicamente racional é uma "emoção racional"; e isso tem a vantagem de nos permitir considerar a calma como um estado emocional, em vez de um padrão privilegiado.

Quando as pessoas pensam em emoção e racionalidade em oposição, suspeito que estejam realmente pensando no Sistema 1 e no Sistema 2 - julgamentos perceptivos rápidos em relação aos julgamentos deliberativos lentos. Julgamentos deliberativos nem sempre são verdadeiros, e julgamentos perceptivos nem sempre são falsos; por isso é muito importante distinguir essa dicotomia de racionalidade. Ambos os sistemas podem servir ao objetivo da verdade, ou derrotá-lo, dependendo de como são usados.

Além da mera curiosidade, existem outros motivos para desejar a verdade? Bem, pode ser que você tenha um objetivo específico no mundo real, como construir um avião e, portanto, precise saber algumas verdades específicas sobre aerodinâmica. Ou talvez algo mais comum, como querer leite com chocolate e, por isso, precisar saber se a mercearia local tem leite com chocolate disponível para decidir se deve ou não ir lá. Se esse for o motivo pelo qual você busca a verdade, então a importância que você dá às suas perguntas refletirá a utilidade que espera obter das informações — o quanto as possíveis respostas afetam suas escolhas, o quanto suas escolhas importam e o quanto você espera que uma resposta diferente possa mudar suas escolhas em relação ao que já é seu padrão.

Buscar a verdade apenas por seu valor instrumental pode parecer impuro — não deveríamos desejar a verdade por si mesma? — mas essas investigações são extremamente importantes porque criam um critério externo de verificação: se o avião cair do céu ou se não houver leite com chocolate na loja, é um sinal de que algo foi feito incorretamente. Você recebe feedback sobre quais modos de pensar funcionam e quais não. A curiosidade pura é algo maravilhoso, mas pode não durar muito depois que o mistério desaparecer. A curiosidade, como emoção humana, existe desde antes dos antigos gregos. Mas o que colocou a humanidade firmemente no Caminho da Ciência foi perceber que certos modos de pensar revelavam crenças que nos permitiam manipular o mundo. Com relação à curiosidade pura, contos giratórios de deuses e heróis em fogueiras satisfaziam esse desejo igualmente e ninguém percebia haver algo de errado com isso.

Existem razões para buscar a verdade além da curiosidade e do pragmatismo? A terceira razão que me vem à mente é a moralidade: acreditar que buscar a verdade é nobre, importante e valioso. Embora esse ideal também atribua um valor intrínseco à verdade, é um estado mental muito diferente da curiosidade. Ser curioso sobre o que está por trás da cortina não é o mesmo que acreditar que você tem o dever moral de olhar para lá. No último estado de espírito, é muito mais provável que você acredite que outra pessoa também deva olhar por trás da cortina ou ser punida se fechar deliberadamente os olhos. Por essa razão, eu também rotularia como "moralidade" a crença de que a busca pela verdade é pragmaticamente importante para a sociedade e, portanto, é um dever de todos. Suas prioridades, sob essa motivação, serão determinadas por seus ideais sobre quais verdades são mais importantes (não mais úteis ou mais intrigantes) ou sobre quando, sob quais circunstâncias, o dever de buscar a verdade é mais forte.

Eu geralmente desconfio da moralidade como uma motivação para a racionalidade, não porque eu rejeite o ideal moral, mas porque isso pode levar a certos tipos de problemas. É muito fácil adquirir modos de pensar que são terríveis erros, na prática, ao aprendermos como deveres morais. Considere o Sr. Spock de Jornada nas Estrelas, um arquétipo ingênuo de racionalidade. Seu estado emocional é sempre definido como "calmo", mesmo quando é completamente inapropriado. Ele geralmente dá muitas casas decimais para probabilidades, grosseiramente descalibradas. Por exemplo: "Capitão, se você conduzir a Enterprise diretamente para aquele buraco negro, nossa probabilidade de sobreviver é de apenas 2,234%". No entanto, nove em cada dez vezes a Enterprise não é destruída. Que tipo de tolo trágico dá quatro casas decimais para uma figura, que está errada em duas ordens de grandeza? No entanto, essa imagem popular é como muitas pessoas concebem o dever de ser "racional", então não é de admirar que elas não aceitem a racionalidade de todo o coração. Transformar a racionalidade em um dever moral é dar a ela todos os terríveis graus de liberdade de um costume tribal arbitrário. As pessoas chegam à resposta errada e depois protestam indignadas que agiram com propriedade, em vez de aprender com o erro.

No entanto, se quisermos melhorar nossas habilidades de racionalidade, ir além dos padrões de desempenho estabelecidos pelos caçadores-coletores, precisaremos de crenças deliberadas sobre como pensar com propriedade. Quando escrevemos novos programas mentais para nós mesmos, eles começam no Sistema 2, o sistema deliberado, e são lentamente — se é que alguma vez — treinados no circuito neural subjacente ao Sistema 1. Então, se há certos tipos de pensamento que queremos evitar — como, digamos, preconceitos — isso acabará sendo representado, no Sistema 2, como uma injunção para não pensar dessa maneira; um dever declarado de evitação.

Se quisermos a verdade, podemos obtê-la de maneira mais eficaz pensando de certas maneiras, em vez de outras; essas são as técnicas da racionalidade. E algumas das técnicas de racionalidade envolvem superar uma certa classe de obstáculos, os preconceitos. . .

4 — … O que é um viés, mesmo?

Um viés é um obstáculo que impede que alcancemos a verdade, uma vez que a busca pela verdade é o nosso objetivo principal. No entanto, muitos obstáculos que não são vieses também podem surgir em nosso caminho.

Começar perguntando "O que é viés?" é começar na ordem errada. Como o provérbio diz: "Existem quarenta tipos de loucura, mas apenas um tipo de bom senso". Alcançar a verdade é difícil, porque é como tentar acertar um alvo muito pequeno em um espaço de configuração muito grande. "Ela me ama, ela não me ama" pode ser uma questão binária, mas $E = mc^2$ é um pontinho no espaço de todas as equações, como um bilhete de loteria premiado no espaço de todos os bilhetes de loteria. Encontrar a verdade é como encontrar um bilhete premiado na loteria, um evento muito improvável que requer uma explicação.

Não temos uma obrigação moral de reduzir vieses, pois os vieses não são inerentemente maus. Acreditar nisso é uma forma de pensar que advém de um senso de dever moral vindo da pressão social. Isso leva as pessoas a tentar aplicar técnicas sem entender a razão por trás delas. Esse tipo de pensamento é visto em *Surely you're joking, Mr. Feynman* (Certamente você está brincando, Sr. Feynman), que li na infância.

Em vez disso, devemos buscar a verdade por qualquer motivo, mesmo que nos deparemos com vários obstáculos no caminho. Esses obstáculos não são completamente diferentes uns dos outros por existirem obstáculos que surgem devido à falta de capacidade de computação ou alto custo da informação. Porém, há um grande grupo de obstáculos que se agrupam em uma região do espaço de obstáculos à verdade, e esse agrupamento é chamado de viés.

O que é um viés? Podemos analisar o agrupamento empírico e encontrar um teste compacto para identificar o que é pertinente? Talvez não seja possível dar uma explicação melhor do que apontar exemplos extensos e esperar que o ouvinte entenda. Se um cientista estiver investigando o fogo pela primeira vez, seria mais eficaz apontar para uma fogueira e dizer: "O fogo é aquela coisa brilhante e alaranjada ali", em vez de dizer: "Eu defino fogo como a transmutação alquímica de substâncias que libera flogisto." Não devemos ignorar algo simplesmente porque não conseguimos defini-lo. Mesmo que eu não consiga citar a fórmula da Relatividade Geral de memória, se eu cair de um penhasco, cairei. O mesmo acontece com os vieses: eles afetam-nos da mesma maneira, independentemente de conseguirmos definir sucintamente o que são.

Com tudo o que foi dito, rotulamos como viés aqueles obstáculos à verdade que são produzidos não pelo custo da informação nem pelo poder computacional limitado, mas pela forma da nossa própria maquinaria mental. Talvez a maquinaria esteja evolutivamente otimizada para propósitos que se opõem ativamente à precisão epistêmica, como a maquinaria para ganhar argumentos em contextos políticos adaptativos. Ou a pressão de seleção foi desviada para a precisão epistêmica, como acreditar no que os outros acreditam para se dar bem socialmente. Ainda, na clássica heurística e viés, a maquinaria opera por um algoritmo identificável que faz algum trabalho útil, mas também produz erros sistemáticos, como a heurística de disponibilidade, que não é uma tendência em si, mas dá origem a tendências identificáveis e descritíveis sucintamente. Nossos cérebros estão fazendo algo errado e, após muita experimentação e/ou pensamento pesado, alguém identifica o problema de uma forma que o Sistema 2 pode compreender; então, chamamos isso de viés. Mesmo que não possamos fazer melhor para conhecer, ainda é uma falha identificável que surge de um tipo particular de maquinário cognitivo — não por ter muito pouco maquinário, mas pela forma do maquinário.

Os vieses são diferentes dos erros que surgem do conteúdo cognitivo, como crenças ou deveres morais adotados. A estes chamamos de erros, em vez de vieses, e são muito mais fáceis de corrigir, uma vez que os notamos por nós mesmos. (Embora a fonte do erro, ou a fonte da fonte do erro, possa, em última análise, ser algum viés.)

Vieses são diferentes de erros que surgem de danos a um cérebro humano individual ou de costumes culturais absorvidos; os preconceitos surgem de uma maquinaria que é universalmente humana.

Platão não era tendencioso por desconhecer a Relatividade Geral — a informação era inacessível para ele; sua ignorância não era um defeito inerente ao seu pensamento. Contudo, se Platão acreditava que filósofos seriam melhores reis por ele próprio ser filósofo — e essa crença se originava de um instinto político de autopromoção, universal e adaptativo, e não porque seu pai lhe dissera que era dever moral de todos promover sua profissão ao governo, ou porque Platão inalava cola na infância —, então isso era um viés, independentemente de Platão ter consciência disso ou não.

Vieses podem não ser baratos para corrigir. Eles podem até não ser corrigíveis. Mas, onde olhamos para nossa própria maquinaria mental e vemos um relato causal de uma classe identificável de erros, e quando o problema parece vir da forma evoluída do maquinário, em vez de haver muito pouco maquinário ou conteúdo específico ruim, então chamamos isso de viés.

Pessoalmente, vejo nossa busca em termos de aquisição de habilidades pessoais de racionalidade e no aperfeiçoamento da técnica de busca da verdade. O desafio é alcançar o objetivo positivo da verdade, não evitar o objetivo negativo do fracasso. O espaço para falhas é amplo, com erros infinitos em variedade infinita. É difícil descrever um espaço tão grande: o que é verdade para uma maçã pode não ser verdade para outra maçã; assim, mais pode ser dito sobre uma única maçã do que sobre todas as maçãs do mundo. O espaço para o sucesso é mais restrito e, portanto, mais pode ser dito sobre ele.

Embora eu não evite (como você pode ver) discutir definições, devemos lembrar que esse não é nosso objetivo principal. Estamos aqui para perseguir a grande busca humana pela verdade: precisamos desesperadamente do conhecimento e, além disso, somos curiosos. Para esse fim, esforcemo-nos para superar quaisquer obstáculos que surjam em nosso caminho, independentemente de chamá-los de vieses ou não.

5 — Disponibilidade

A heurística da disponibilidade consiste em julgar a frequência ou probabilidade de um evento com base na facilidade com que exemplos desse evento vêm à mente.

Um estudo famoso de 1978 realizado por Lichtenstein, Slovic, Fischhoff, Layman e Combs, intitulado *Judged Frequency of Lethal Events* (Frequência estimada de eventos letais), analisou erros na quantificação da gravidade dos riscos, ou seja, na avaliação de qual de dois perigos ocorria com mais frequência. Os participantes acreditavam que acidentes causavam tantas mortes quanto as doenças e, que o homicídio era mais comum do que o suicídio. Na verdade, as doenças causam cerca de dezesseis vezes mais mortes do que os acidentes, e o suicídio é duas vezes mais comum do que o homicídio.

Uma explicação óbvia para essas crenças distorcidas é ser mais provável, que as pessoas comentem sobre assassinatos do que sobre suicídios, facilitando que alguém se lembre de ter ouvido falar de um assassinato do que de um suicídio. Acidentes são mais dramáticos do que doenças, o que pode tornar as pessoas mais propensas a se lembrar de um acidente. Em 1979, um estudo subsequente de Combs e Slovic mostrou que julgamentos distorcidos de probabilidade se correlacionavam fortemente (0,85 e 0,89) com frequências distorcidas de reportagens em dois jornais. Isso não esclarece se os assassinatos estão mais disponíveis para a memória porque são mais relatados, ou se os jornais informam mais sobre assassinatos porque eles são mais vívidos (e, portanto, mais lembrados). De qualquer forma, um viés de disponibilidade está em ação.

O relato seletivo é uma das principais fontes de viés de disponibilidade. No ambiente ancestral, muito do que se conhecia era vivenciado diretamente ou ouvido diretamente de um membro da tribo que testemunhou o evento. Normalmente, havia apenas uma camada de relato seletivo entre você e o evento em si. Hoje, com a Internet, é possível ver relatos que passaram pelas mãos de seis blogueiros antes de chegar até você, ou seja, seis filtros sucessivos. Em comparação com nossos ancestrais, vivemos em um mundo maior, onde muito mais acontece e muito menos chega até nós — um efeito de seleção muito mais forte, que pode criar vieses de disponibilidade muito maiores.

Na realidade, é improvável que se conheça Bill Gates pessoalmente. No entanto, graças aos relatos seletivos da mídia, pode-se ficar tentado a comparar o sucesso da própria vida com o dele e sofrer penalidades hedônicas em consequência. A frequência objetiva de Bill Gates é de 0,00000000015, mas se ouve falar dele com muito mais frequência. Por outro lado, 19% do planeta vive com menos de US$ 1 por dia, e é duvidoso que um quinto dos posts de blog que se leem sejam escritos por eles.

Usar a disponibilidade pode gerar um viés de absurdo, pois eventos que nunca aconteceram são esquecidos e, portanto, considerados tendo probabilidade zero. Quando não há histórico recente de inundações (e mesmo assim as probabilidades ainda podem ser razoavelmente calculadas), as pessoas se recusam a comprar seguro contra inundações, mesmo quando o prêmio é fortemente subsidiado e o preço é muito abaixo do valor atuarial justo. Kunreuther e outros sugerem que a reação insuficiente às ameaças de inundação pode ser resultado da "incapacidade das pessoas em conceituar inundações que nunca ocorreram… As pessoas nas planícies de inundação parecem prisioneiras de sua experiência… Inundações recentes parecem estabelecer um limite superior para o tamanho da perda com o qual os gestores acreditam que devem se preocupar"

Burton e outros relatam que, quando barragens e diques são construídos, eles reduzem a frequência das inundações e, assim, criam uma falsa sensação de segurança, levando a precauções reduzidas, enquanto dano médio anual aumenta. Os sábios extrapolarão a partir da memória de pequenos perigos para a possibilidade de grandes perigos. No entanto, a experiência de pequenos perigos parece definir um limite percebido para o risco. Uma sociedade bem protegida contra perigos menores não toma nenhuma ação contra riscos maiores, construindo em áreas sujeitas a inundações, uma vez que as inundações menores regulares são eliminadas. Uma sociedade sujeita a perigos menores regulares trata esses perigos menores como um limite superior no tamanho dos riscos, protegendo-se contra pequenas inundações regulares, mas não contra grandes inundações ocasionais.

A memória nem sempre é um bom guia para probabilidades no passado, muito menos no futuro.

Referências

1. *Sarah Lichtenstein et al., "Judged Frequency of Lethal Events," Journal of Experimental Psychology: Human Learning and Memory 4, no. 6 (1978): 551–578, doi:10.1037/0278-7393.4.6.551.*
2. *Barbara Combs and Paul Slovic, "Newspaper Coverage of Causes of Death," Journalism & Mass Communication Quarterly 56, no. 4 (1979): 837–849, doi:10.1177/107769907905600420.*
3. *Howard Kunreuther, Robin Hogarth, and Jacqueline Meszaros, "Insurer Ambiguity and Market Failure," Journal of Risk and Uncertainty 7 (1 1993): 71–87, doi:10.1007/BF01065315.*
4. *Ian Burton, Robert W. Kates, and Gilbert F. White, The Environment as Hazard, 1st ed. (New York: Oxford University Press, 1978).*

6 — Detalhes onerosos

"Meramente um detalhe corroborativo, destinado a dar verossimilhança artística a uma narrativa vazia e pouco convincente"[3].

— Pooh-Bah, em The Mikado (O Mikado) de Gilbert e Sullivan

A falácia da conjunção ocorre quando as pessoas avaliam a probabilidade P(A, B) como sendo maior do que a probabilidade P(B), embora um teorema mostre que P(A, B) ≤ P(B). Por exemplo, em um experimento realizado em 1981, 68% dos participantes classificaram como mais provável a afirmação "Reagan fornecerá apoio federal para mães solteiras e cortará o apoio federal aos governos locais" do que a afirmação "Reagan fornecerá apoio federal para mães solteiras"

Uma série de experimentos habilmente planejados, que eliminaram hipóteses alternativas e definiram a interpretação padrão, confirmaram que a falácia da conjunção ocorre porque "substituímos o julgamento de representatividade pelo julgamento de probabilidade". Ao adicionar detalhes extras, você pode fazer com que um resultado pareça mais característico do que o processo que o gerou. Você pode tornar mais plausível que Reagan apoiará mães solteiras, acrescentando a alegação de que Reagan também cortará o apoio aos governos locais. A implausibilidade de uma afirmação é compensada pela plausibilidade da outra, elas "ficam na média".

O que isso significa é que adicionar detalhes pode fazer com que um cenário pareça mais plausível, mesmo que o evento necessariamente se torne menos provável.

Se for assim, então hipoteticamente, podemos encontrar futuristas inventando histórias inescrupulosamente plausíveis e detalhadas, ou encontrar pessoas engolindo enormes pacotes de reivindicações sem suporte, agrupadas com algumas afirmações de som forte no centro. Se você se deparar com a falácia da conjunção em uma comparação direta, poderá ter sucesso nesse problema específico corrigindo-se conscientemente. Mas isso é apenas colocar um curativo no problema, não o corrigir, em geral.

No experimento de 1982, em que previsores profissionais atribuíram probabilidades mais altas à afirmação "Rússia invade a Polônia, seguida de suspensão das relações diplomáticas entre os EUA e a URSS" do que à afirmação "Suspensão das relações diplomáticas entre os EUA e a URSS", cada grupo experimental foi apresentado com uma proposta. Que estratégia esses previsores poderiam ter seguido, como um grupo, para eliminar a falácia da conjunção, quando ninguém sabia exatamente sobre a comparação? Quando nenhum indivíduo sequer sabia que o experimento era sobre a falácia da conjunção? Como eles poderiam ter se saído melhor em seus julgamentos de probabilidade?

3 NT: Tradução livre do texto original em inglês. *Merely corroborative detail, intended to give artistic verisimilitude to an otherwise bald and unconvincing narrative. . .*

Corrigir uma pegadinha como um caso especial não resolve o problema geral. A pegadinha é o sintoma, não, a doença.

O que poderiam ter feito os previsores para evitar a falácia da conjunção, sem ver a comparação direta, ou mesmo sabendo que alguém iria testá-los na falácia da conjunção? Parece-me que eles precisariam notar a palavra "e". Eles teriam que ser cautelosos — não apenas cautelosos, mas recuar. Mesmo sem saber que os pesquisadores iriam testá-los mais tarde na falácia da conjunção em particular, eles deveriam notar a junção de dois detalhes inteiros e ficariam chocados com a audácia de qualquer um que lhes pedisse para endossar uma previsão tão insanamente complicada. Eles teriam que penalizar substancialmente a probabilidade — um fator de quatro, pelo menos, conforme os detalhes experimentais.

Também pode ter ajudado os previsores a pensar sobre as possíveis razões pelas quais os EUA e a União Soviética suspenderiam as relações diplomáticas. O cenário não é "Os EUA e a União Soviética suspendem repentinamente as relações diplomáticas sem motivo", mas "Os EUA e a União Soviética suspendem as relações diplomáticas por qualquer motivo".

E os sujeitos que avaliaram "Reagan fornecerá apoio federal para mães solteiras e cortará o apoio federal aos governos locais"? Novamente, eles precisariam ficar chocados com a palavra "e". Além disso, eles deveriam adicionar absurdos — onde o absurdo é o logaritmo da probabilidade, então você pode adicioná-lo — em vez de tirar a média deles. Eles deveriam pensar: "Reagan pode ou não cortar o apoio aos governos locais (1 bit), mas parece muito improvável que ele apoie mães solteiras (4 bits). Total de absurdos: 5 bits." Ou talvez: "Reagan não apoiará mães solteiras. Um golpe e acabou. A outra proposição só torna tudo ainda pior".

Da mesma forma, considere o dado de seis faces com quatro faces azuis e duas vermelhas. Os sujeitos tiveram que apostar na sequência (1) VAVVV, (2) AVAVVV ou (3) AVVVVV aparecendo em qualquer lugar em vinte lançamentos de dados. Sessenta e cinco por cento dos sujeitos escolheram AVAVVV, sendo estritamente dominado por VAVVV, já que qualquer sequência contendo AVAVVV também é válida para VAVVV. Como os sujeitos poderiam ter se saído melhor? Ao perceber a inclusão? Talvez, mas isso é apenas um curativo, não resolve o problema fundamental. Calculando explicitamente as probabilidades? Isso certamente resolveria o problema fundamental, mas nem sempre é possível calcular uma probabilidade exata.

Os sujeitos cometeram um erro heurístico ao pensar: "Ha! A sequência 2 tem a maior proporção de azul para vermelho! Eu deveria apostar na Sequência 2!" Para vencer heuristicamente, os sujeitos precisariam pensar: "Ha! A sequência 1 é mais curta! Eu deveria ir com a Sequência 1!".

Eles precisariam sentir um impacto emocional mais forte da Navalha de Ocam — sentindo cada detalhe adicionado como um fardo, mesmo uma única jogada extra de dados.

Uma vez, eu estava conversando com alguém que havia sido hipnotizado por um futurista enganoso (aquele que adiciona muitos detalhes que parecem impressionantes). Eu estava tentando explicar por que não estava igualmente hipnotizado por essas teorias incríveis e impressionantes. Então, eu expliquei sobre a falácia da conjunção, especificamente o experimento de "suspender relações ± invadir a Polônia". E ele disse: "Ok, mas o que isso tem a ver com…". Eu disse: "É mais provável que os universos se repliquem por qualquer motivo, do que eles se repliquem por meio de buracos negros porque as civilizações avançadas fabricam buracos negros porque os universos evoluem para fazê-los fazer isso". E ele disse: "Ah".

Até então, ele não sentia esses detalhes extras como fardos adicionais. Em vez disso, eram detalhes corroborativos, conferindo verossimilhança à narrativa. Alguém apresenta a você um pacote de ideias estranhas, uma das quais é que os universos se replicam. Em seguida, eles apresentam suporte para a afirmação de que os universos se replicam. Mas isso não é suporte para o pacote, embora tudo seja contado como uma história.

Você tem que separar os detalhes. Você deve levantar cada um independentemente e perguntar: "Como sabemos esse detalhe?" Alguém desenha uma imagem da queda da humanidade na guerra nanotecnológica, onde a China se recusa a cumprir um acordo de controle internacional, seguido por uma corrida armamentista… Espere um minuto — como você sabe que será a China? Você tem uma bola de cristal no seu bolso ou está feliz em ser um futurista? De onde vêm todos esses detalhes? De onde veio esse detalhe específico?

Pois está escrito:

Se você pode aliviar seu fardo, deve fazê-lo.

Até a menor gota d'água pode fazer o copo transbordar…

Referências

1. *William S. Gilbert and Arthur Sullivan, The Mikado, Opera, 1885.*
2. *Tversky and Kahneman, "Extensional Versus Intuitive Reasoning."*
3. *Amos Tversky and Daniel Kahneman, "Judgments of and by Representativeness," in Judgment Under Uncertainty: Heuristics and Biases, ed. Daniel Kahneman, Paul Slovic, and Amos Tversky (New York: Cambridge University Press, 1982), 84–98.*

7 — Falácia do planejamento

O Aeroporto Internacional de Denver foi inaugurado com um atraso de 16 meses, a um custo que excedeu US$ 2 bilhões (também foi declarado um valor de US$ 3,1 bilhões). O Eurofighter Typhoon, um projeto de defesa conjunto de vários países europeus, foi entregue com um atraso de 54 meses a um custo de US$ 19 bilhões, em vez de US$ 7 bilhões. A Ópera de Sydney pode ser a construção mais lendária de todos os tempos, originalmente estimada para ser concluída em 1963 por US$ 7 milhões e finalmente concluída em 1973 por US$ 102 milhões.[1]

Esses desastres isolados são trazidos à nossa atenção pela disponibilidade seletiva? São sintomas de burocracia ou falhas de incentivo do governo? Muito provavelmente, sim. Mas há também um viés cognitivo correspondente, replicado em experimentos com planejadores individuais.

Bühler e outros, pediram a seus alunos estimativas de quando eles (os alunos) pensavam que concluiriam seus projetos acadêmicos pessoais. Mais especificamente, os pesquisadores solicitaram aos alunos que definissem prazos para os quais eles acreditavam ter 50%, 75% e 99% de probabilidade de concluir seus projetos. Você consegue adivinhar quantos alunos terminaram seus projetos antes ou nos prazos estimados de 50%, 75% e 99%?

- 13% dos participantes terminaram seu projeto antes de terem atribuído um nível de probabilidade de 50%;
- 19% terminaram no tempo atribuído a um nível de probabilidade de 75%;
- E apenas 45% (menos da metade!) terminaram no momento de seu nível de probabilidade de 99%.

Como Buehler e outros escreveram:

“Os resultados para o nível de probabilidade de 99% são especialmente impressionantes: Mesmo quando solicitados a fazer uma previsão altamente conservadora, uma previsão que eles tinham quase certeza de que cumpririam, a confiança dos alunos em suas estimativas de tempo superou em muito suas realizações.”[3]

De forma geral, esse fenômeno é conhecido como “falácia do planejamento”. Essa falácia é que as pessoas acreditam que podem planejar, risos.

Newby-Clark e sua equipe descobriram uma pista para o problema subjacente ao algoritmo de planejamento. Eles descobriram que:

- Pedir aos participantes suas previsões com base em cenários realistas de melhor palpite; e
- Perguntar aos participantes sobre seus cenários de melhor caso esperados… produziu resultados indistinguíveis.[4]

Quando as pessoas são solicitadas a apresentar um cenário realista, elas visualizam tudo como exatamente planejado, sem atrasos inesperados ou catástrofes imprevistas — a mesma visão de seu “melhor cenário”.

A realidade muitas vezes oferece resultados um pouco piores do que o pior caso.

Ao contrário da maioria dos vieses cognitivos, conhecemos uma boa heurística para eliminar vieses na falácia do planejamento. Ela pode não funcionar para caos em grande escala, como o Aeroporto Internacional de Denver, mas funciona bem para planejamentos pessoais e até mesmo para algumas questões organizacionais de menor escala. Tudo o que é necessário é usar uma visão externa em vez de uma visão interna.

As pessoas tendem a criar suas previsões pensando nas características específicas e únicas da tarefa em questão e criando um cenário de como pretendem concluir a tarefa — o que chamamos de planejamento.

Quando você deseja fazer algo, precisa planejar onde, quando e como; descobrir quanto tempo e recursos são necessários; visualizar as etapas desde o início até a conclusão bem-sucedida. Tudo isso é uma visão interna e não considera atrasos inesperados ou catástrofes imprevistas. Como já foi mencionado, pedir às pessoas que visualizem o pior caso ainda não é suficiente para neutralizar seu otimismo — elas não visualizam o suficiente o potencial para problemas.

A visão externa é quando você deliberadamente evita pensar nas características especiais e únicas deste projeto e simplesmente pergunta quanto tempo levou para concluir projetos amplamente semelhantes no passado. Isso é contra-intuitivo, já que a visão interna tem muito mais detalhes — há uma tentação de pensar que uma previsão cuidadosamente adaptada, considerando todos os dados disponíveis, dará melhores resultados.

No entanto, a experiência mostrou que quanto mais detalhada a visualização das pessoas, mais otimistas (e menos precisas) elas se tornam. Em um estudo, Buhler e outros pediram a um grupo experimental de participantes que descrevesse planos altamente específicos para suas compras de Natal — onde, quando e como. Em média, esse grupo esperava terminar as compras mais de uma semana antes do Natal. Outro grupo foi simplesmente perguntado quando esperava terminar suas compras de Natal, com uma resposta média de quatro dias. Ambos os grupos terminaram em média três dias antes do Natal.

Da mesma forma, Buehler et al., relatando um estudo transcultural, descobriram que os estudantes japoneses esperavam terminar suas redações dez dias antes do prazo. Na verdade, eles terminaram um dia antes do prazo. Quando perguntados quando já haviam concluído tarefas semelhantes, eles responderam: um dia antes do prazo.[6] Essa é a força da visão externa sobre a visão interna.

Outra descoberta semelhante é que pessoas experientes de fora, que conhecem menos os detalhes, mas têm memória relevante para se basear, geralmente são muito menos otimistas e muito mais precisas do que os planejadores e implementadores reais.

Assim, para corrigir a falácia do planejamento, existe uma maneira confiável quando você está realizando algo que é amplamente semelhante a uma classe de referência de projetos anteriores. Você só precisa perguntar quanto tempo os projetos semelhantes levaram no passado, sem considerar nenhuma das propriedades especiais deste projeto. É ainda melhor perguntar a um estranho experiente quanto tempo os projetos semelhantes levaram.

Ao fazer isso, você receberá uma resposta que pode parecer extremamente longa e refletir claramente a falta de compreensão dos motivos especiais pelos quais esta tarefa em particular levará menos tempo. No entanto, essa resposta é verdadeira e deve ser aceita. Lide com isso.

Referências

1. *Roger Buehler, Dale Griffin, and Michael Ross, "Inside the Planning Fallacy: The Causes and Consequences of Optimistic Time Predictions," em Gilovich, Griffin, e Kahneman, Heuristics and Biases, 250–270.*
2. *Roger Buehler, Dale Griffin, and Michael Ross, "Exploring the 'Planning Fallacy': Why People Underestimate Their Task Completion Times," Journal of Personality and Social Psychology 67, no. 3 (1994): 366–381, doi:10.1037/0022-3514.67.3.366; Roger Buehler, Dale Griffin, and Michael Ross, "It's About Time: Optimistic Predictions in Work and Love," European Review of Social Psychology 6, no. 1 (1995): 1–32, doi:10.1080/14792779343000112.*
3. *Buehler, Griffin, and Ross, "Inside the Planning Fallacy."*
4. *Ian R. Newby-Clark et al., "People Focus on Optimistic Scenarios and Disregard Pessimistic Scenarios While Predicting Task Completion Times," Journal of Experimental Psychology: Applied 6, no. 3 (2000): 171–182, doi:10.1037/1076-898X.6.3.171.*
5. *Buehler, Griffin, and Ross, "Inside the Planning Fallacy."*
6. *Ibid.*

8 — A ilusão de transparência: por que ninguém te entende

No viés retrospectivo, as pessoas que conhecem o resultado de uma situação acreditam que o resultado deveria ter sido fácil de prever com antecedência. Conhecendo o resultado, re-interpretamos a situação à luz desse resultado. Mesmo quando avisados, não podemos "desinterpretar" para simpatizar com alguém que não sabe o que sabemos.

A ilusão de transparência está intimamente ligada a esse viés. Sempre sabemos o que queremos dizer com nossas palavras e esperamos que os outros também saibam. Ao ler nossa própria escrita, a interpretação pretendida se encaixa facilmente, guiada pelo nosso conhecimento do que realmente queremos dizer. É difícil se colocar no lugar de alguém que precisa interpretar nossas palavras sem ter o mesmo contexto ou conhecimento.

June recomenda um restaurante para Mark; Mark janta lá e descobre uma (a) comida inexpressiva e um serviço medíocre ou (b) uma comida deliciosa e um serviço impecável. Então, Mark deixa a seguinte mensagem na secretária eletrônica de June: "June, acabei de jantar no restaurante que você recomendou e devo dizer que foi maravilhoso, simplesmente maravilhoso." Keysar apresentou a um grupo de sujeitos o cenário (a), e 59% acharam que a mensagem de Mark era sarcástica e Jane perceberia o sarcasmo. [1] Entre outros sujeitos, contados do cenário (b), apenas 3% pensaram que Jane perceberia a mensagem de Mark tão sarcástico. Keysar e Barr parecem indicar que uma mensagem de voz real foi reproduzida para os sujeitos. [2] Keysar mostrou que se os sujeitos fossem informados de que o restaurante era horrível, mas que Mark queria esconder sua resposta, eles acreditavam que June não perceberia sarcasmo na (mesma) mensagem: [3]

Eles eram tão propensos a prever que ela perceberia sarcasmo quando ele tentasse esconder sua experiência negativa quanto quando ele tivesse uma experiência positiva e fosse verdadeiramente sincero. Assim, os participantes consideraram transparente a intenção comunicativa de Mark. Era como se presumissem que June perceberia qualquer intenção que Mark quisesse que ela percebesse. [4]

"The goose hangs high" é uma expressão arcaica do inglês que está em desuso na linguagem moderna. Keysar e Barr disseram a um grupo de participantes que "o ganso está pendurado" significava que o futuro parecia bom; outro grupo de sujeitos aprendeu que "o ganso está pendurado" significava que o futuro parece sombrio. [5] Os sujeitos foram então questionados sobre qual desses dois significados um ouvinte desinformado teria mais probabilidade de atribuir ao idioma. Cada grupo pensou que os ouvintes perceberiam o significado apresentado como "padrão".

(Outras expressões idiomáticas testadas incluíam "come the uncle over someone," "to go by the board," e "to lay out in lavender." Ah, inglês, uma língua tão adorável.) Keysar e Henly testaram a calibração dos falantes: os falantes subestimariam, superestimariam ou estimariam corretamente a frequência com que os ouvintes os entendiam? [6] Os falantes recebiam sentenças ambíguas ("O homem está perseguindo uma mulher em uma bicicleta."), em seguida, pediu aos falantes que pronunciassem as palavras na frente dos ouvintes e, em seguida, pediu aos falantes que estimassem quantos ouvintes entenderam o significado pretendido. Os falantes pensaram que foram compreendidos em 72% dos casos e foram realmente compreendidos em 61% dos casos. Quando os ouvintes não entenderam, os falantes pensaram que entenderam em 46% dos casos; quando os ouvintes entenderam, os falantes pensaram que não em apenas 12% dos casos.

Participantes adicionais que ouviram a explicação não mostraram tal viés, presumindo que os ouvintes compreenderiam em apenas 56% dos casos.

Keysar e Barr observaram que, dois dias antes do ataque da Alemanha à Polônia, Chamberlain enviou uma carta com a intenção de deixar claro que a Grã-Bretanha lutaria se ocorresse alguma invasão. [7] A carta, redigida de maneira polida e diplomática, foi interpretada por Hitler como conciliatória e, consequentemente, os tanques alemães avançaram.

Não se apresse em culpar aqueles que interpretam mal suas frases perfeitamente claras, faladas ou escritas. As chances são de que suas palavras sejam mais ambíguas do que você pensa.

Referências

1. *Boaz Keysar, "The Illusory Transparency of Intention: Linguistic Perspective Taking in Text,"Cognitive Psychology 26 (2 1994): 165–208, doi:10.1006/cogp.1994.1006.*
2. *Keysar and Barr, "Self-Anchoring in Conversation."*
3. *Boaz Keysar, "Language Users as Problem Solvers: Just What Ambiguity Problem Do They Solve?," in Social and Cognitive Approaches to Interpersonal Communication, ed. Susan R. Fussell and Roger J. Kreuz (Mahwah, NJ: Lawrence Erlbaum Associates, 1998), 175–200.*
4. *Keysar and Barr, "Self-Anchoring in Conversation."*
5. *Boaz Keysar and Bridget Bly, "Intuitions of the Transparency of Idioms: Can One Keep a Secret by Spilling the Beans?," Journal of Memory and Language 34 (1 1995): 89–109, doi:10.1006/jmla.1995.1005.*
6. *Boaz Keysar and Anne S. Henly, "Speakers' Overestimation of Their Effectiveness," Psychological* Science 13 (3 2002): 207–212, doi:10.1111/1467-9280.00439.
7. Keysar and Barr, "Self-Anchoring in Conversation."

9 — Presumindo distâncias inferenciais curtas

O ambiente de adaptação evolutiva do Homo sapiens (também conhecido como EEA ou "ambiente ancestral") consistia em bandos de caçadores-coletores de no máximo 200 pessoas, sem escrita. Todo o conhecimento herdado foi transmitido pela fala e pela memória. Nesse mundo, todo conhecimento prévio é considerado conhecimento universal. Todas as informações que não são estritamente privadas são públicas, ponto final.

No ambiente ancestral, era improvável que você terminasse a mais de um passo inferencial de qualquer outra pessoa. Quando você descobre um novo oásis, não precisa explicar aos seus companheiros de tribo o que é um oásis, ou por que é uma boa ideia beber água, ou como caminhar. Só você sabe onde fica o oásis; isso é conhecimento privado. Mas todos têm o background para entender sua descrição do oásis, os conceitos necessários para pensar sobre a água; isso é conhecimento universal. Quando você explica as coisas em um ambiente ancestral, raramente precisa explicar seus conceitos. No máximo, você tem que explicar um novo conceito, não dois ou mais simultaneamente.

No contexto ancestral, não havia disciplinas abstratas com vastos corpos de evidências cuidadosamente reunidas generalizadas em teorias elegantes transmitidas por livros escritos cujas conclusões eram cem passos inferenciais de distância das premissas de fundo universalmente compartilhadas.

No ambiente dos nossos ancestrais, quem dizia algo sem apoio aparente era considerado um mentiroso ou um idiota. Era improvável que você pensasse: "Ei, talvez essa pessoa tenha um conhecimento de fundo bem fundamentado do qual ninguém na minha tribo sequer ouviu falar", porque isso não aconteceu devido a uma invariante confiável do ambiente ancestral.

Por outro lado, se você dizia algo obviamente óbvio e a outra pessoa não percebesse, ela é a idiota ou está sendo deliberadamente obstinada em irritá-lo.

E ainda por cima, se alguém dizia algo sem nenhum apoio óbvio e esperava que você acreditasse, agindo indignado quando você não acreditava, então essa pessoa deveria estar louca.

Combinado com a ilusão de transparência e autoancoragem, acho que isso explica muito sobre a lendária dificuldade que a maioria dos cientistas tem em se comunicar com um público leigo — ou mesmo se comunicar com cientistas de outras disciplinas. Quando observo falhas de explicação, geralmente vejo o explicador dando um passo para trás, quando precisa dar dois ou mais passos para trás. Ou os ouvintes assumem que as coisas devem ser visíveis em uma etapa, quando dão duas ou mais etapas para explicar. Ambos os lados agem como se esperassem distâncias inferenciais muito curtas do conhecimento universal para qualquer novo conhecimento.

Um biólogo, falando com um físico, pode justificar a evolução dizendo ser a explicação mais simples. Mas nem todos na Terra foram inculcados com aquela lendária história da ciência, de Newton a Einstein, que investe a frase "explicação mais simples" com seu significado impressionante: uma Palavra de Poder, falada no nascimento de teorias e esculpida em suas lápides. Para outra pessoa: "Mas é a explicação mais simples!" pode soar como um argumento interessante, mas dificilmente arrasador; não parece uma ferramenta tão poderosa para compreender a política do escritório ou consertar um carro quebrado. Obviamente, o biólogo está apaixonado por suas próprias ideias, muito arrogante para estar aberto a explicações alternativas que soam igualmente plausíveis. (Se isso soa plausível para mim, deve soar plausível para qualquer membro da minha banda.)

Do ponto de vista do biólogo, é possível entender como a evolução pode parecer um pouco estranha no começo. Mas quando alguém rejeita a evolução mesmo depois que o biólogo explica que é a explicação mais simples, bem, é evidente que não-cientistas são apenas idiotas e não faz sentido argumentar com eles.

Um argumento claro deve apresentar um raciocínio inferencial que começa com o que o público já sabe ou aceita. Caso contrário, você estará falando sozinho.

Se em algum momento você fizer uma declaração sem uma justificativa clara com base nos argumentos que você defendeu anteriormente, o público simplesmente pensará que você está louco.

Isso também ocorre quando você é percebido atribuindo um peso maior a um argumento do que é justificado na perspectiva do público naquele momento. Por exemplo, falar como se acreditar que a "explicação mais simples" é um argumento que arrasa a evolução (o que é verdade), em vez de uma ideia apenas interessante (como pode parecer para alguém que não foi ensinado a valorizar a Navalha de Ocam).

Ah, e é melhor você não dar nenhuma dica de que acha que está trabalhando dúzia de passos inferenciais do que o público sabe, ou que você acha que tem um conhecimento prévio especial não disponível para eles. O público não sabe nada sobre um argumento psicológico evolutivo para um viés cognitivo para subestimar as distâncias inferenciais que levam a engarrafamentos na comunicação. Eles só vão pensar que você é condescendente.

E se você acha que pode explicar o conceito de "distâncias de inferência sistematicamente subestimadas" brevemente, com poucas palavras, infelizmente tenho uma má notícia para você…

10 — A lente que vê suas próprias falhas

Os raios de luz emanam do Sol e atingem seus cadarços e ricocheteiam; alguns fótons entram nas pupilas de seus olhos, atingindo sua retina; a energia dos fótons desencadeia impulsos neurais; os impulsos neurais são transmitidos às áreas de processamento visual do cérebro; onde a informação ótica é processada e reconstruída em um modelo 3D, que é, então, reconhecido por você como um cadarço desamarrado; e assim, você acredita que os seus cadarços estão desamarrados.

Este é o segredo da racionalidade deliberada — todo esse processo não é magia, e você consegue compreendê-lo. Você entende como consegue ver seus cadarços. Você consegue pensar quais tipos de processos mentais darão origem a crenças que espelharão a realidade, e, quais deles não farão o mesmo.

Camundongos enxergam, mas não entendem por que enxergam. Você entende a visão, e por causa disso, você consegue fazer coisas que os camundongos não conseguem. Pare um pouco e se maravilhe com isso, pois isso é realmente algo maravilhoso.

Camundongos veem, mas não sabem que possuem córtex visual, portanto, não conseguem corrigir ilusões de ótica. O universo mental habitado pelo camundongo inclui gatos, buracos, queijo e ratoeiras — mas não cérebros de camundongos. Suas câmeras não tiram fotos de suas próprias lentes. Porém, nós, como seres humanos, podemos olhar para uma imagem aparentemente bizarra, e compreender que parte daquilo que estamos enxergando é a própria lente em si. Você não precisa acreditar sempre naquilo que vê, mas precisa ter consciência de que possui olhos — possui compartimentos mentais distintos para organizar o mapa e o território, os sentidos e a realidade. Antes que você pense que essa habilidade é trivial, lembre-se de que ela é rara no reino animal.

A ideia de ciência é, simplesmente, o raciocínio reflexivo sobre um processo mais confiável para fazer os conteúdos da sua mente espelharem os conteúdos do mundo. É o tipo de coisa que os camundongos jamais inventariam. Ao ponderar sobre esse negócio de "realizar experimentos replicáveis para falsificar teorias," dá para ver por que isso funciona. A ciência não é um magistério à parte, distante da realidade e do entendimento de meros mortais. A ciência não é nada que se aplica somente ao interior dos laboratórios. A ciência, em si, é um processo compreensível que acontece no mundo, e que, correlaciona cérebros com a realidade.

Ela faz sentido, quando pensamos sobre ela. Mas camundongos não conseguem pensar sobre o pensar, e, é por isso que eles não têm ciência. Não devemos ignorar o quanto isso é maravilhoso — ou o poder potencial que isso confere a nós como indivíduos, não apenas como sociedades científicas.

Reconhecidamente, entender a máquina do pensamento pode ser um pouco mais complicado do que entender uma máquina a vapor — mas não é uma tarefa fundamentalmente diferente.

Uma vez, eu entrei numa sala de bate-papo sobre filosofia na EFNet e perguntei: "Vocês acreditam que uma guerra nuclear acontecerá nos próximos 20 anos? Em caso negativo, por que não?" Uma pessoa que respondeu à minha pergunta disse que não esperava que uma guerra nuclear fosse acontecer por 100 anos, porque "Todos os atores envolvidos nas decisões sobre guerras nucleares, não estavam interessados numa guerra agora." "Mas por que prolongar esse prazo para 100 anos?" Perguntei. "Esperança, só isso", ele respondeu.

Ao refletir sobre esse processo de raciocínio, conseguimos entender que a ideia de uma guerra nuclear deixa uma pessoa triste, e entendemos como, consequentemente, seu cérebro rejeita essa crença. Mas se você imaginar um bilhão de mundos — ramificações de Everett ou duplicatas de Tegmark [1], esse processo de pensamento não relacionará sistematicamente os otimistas a ramificações nas quais não ocorre nenhuma guerra nuclear. (Algum sujeito inteligente está fadado a dizer: "Ah, mas já que tenho esperança, vou trabalhar um pouco mais no meu trabalho, impulsionar a economia global e, assim, ajudar a impedir que os países caiam no estado raivoso e sem esperança em que uma guerra nuclear é uma possibilidade. Então, ambos os eventos estão, sim, relacionados, afinal." A essa altura, precisamos arrastar o Teorema de Bayes para a conversa e medir essa relação quantitativamente. Sua natureza otimista não pode ter um efeito tão grande no mundo; não pode, por si só, diminuir a probabilidade de uma guerra nuclear em 20%, ou por mais que sua natureza otimista mude suas crenças. Deslocar muito as suas crenças, devido a um evento que aumenta em muito pouco as chances de você estar correto, ainda assim bagunçará o seu mapeamento.

Perguntar quais das suas crenças te fazem feliz é voltar-se para dentro de si mesmo, e, não, para fora — isso te diz algo sobre si, mas não é uma evidência atrelada ao ambiente externo. Eu não sou contra a felicidade, mas ela deveria acompanhar a sua própria visão de mundo, ao invés de interferir em suas ferramentas mentais.

Se você consegue entender isso — se consegue entender que a esperança está deslocando demais os seus pensamentos de primeira ordem — se reconhecer a sua mente como uma espécie de motor de mapeamento que possui falhas — então você conseguirá aplicar uma correção reflexiva. O cérebro é uma lente defeituosa através da qual enxergamos a realidade. Isso se aplica tanto aos cérebros dos camundongos quanto aos cérebros dos seres humanos. Porém, o cérebro humano é uma lente defeituosa capaz de entender suas próprias falhas — seus erros sistemáticos, seus vieses — e implementar correções de segunda ordem. Isto, na prática, torna nossas lentes muito mais poderosas. Não perfeitas, mas muito mais poderosas.

Referências

1. *Max Tegmark, "Parallel Universes," in Science and Ultimate Reality: Quantum Theory, Cosmology, and Complexity, ed. John D. Barrow, Paul C. W. Davies, and Charles L. Harper Jr. (New York: Cambridge University Press, 2004), 459–491.*

B — Crenças falsas

11 — Fazendo crenças pagarem aluguel (em experiências antecipadas)

Assim começa a antiga parábola:

Se uma árvore cai na floresta e ninguém a ouve, ela faz barulho? Alguém diz: “Sim, faz, ao criar vibrações no ar”. Outro diz: “Não, não faz, pois nenhum cérebro processa os sons”.

Suponha que, após a queda da árvore, os dois caminhem juntos para a floresta. Será que um deles espera ver a árvore caída para a direita e o outro espera ver a árvore caída para a esquerda? Suponha que, antes da árvore cair, os dois deixem um gravador de som ao lado da árvore. Alguém, tocando o gravador, esperaria ouvir algo diferente do outro? Suponha que eles conectem um eletroencefalógrafo a qualquer cérebro do mundo; alguém esperaria ver um traço diferente do outro? Embora ambos discutam, um dizendo “Não” e o outro dizendo “Sim”, eles não antecipam nenhuma experiência diferente. Os dois pensam que têm modelos diferentes do mundo, mas não têm diferença em relação ao que esperam que lhes aconteça.

É tentador tentar eliminar essa classe de erro, insistindo que o único tipo legítimo de crença é uma antecipação da experiência sensorial. Mas o mundo contém, de fato, muito do que não é percebido diretamente. Não vemos os átomos subjacentes ao tijolo, mas os átomos de fato estão lá. Há um chão sob seus pés, mas você não vivencia o chão diretamente; você vê a luz refletida do chão, ou melhor, você vê o que sua retina e córtex visual processaram dessa luz. Inferir o chão a partir da visão do chão é retroceder nas causas invisíveis da experiência. Pode parecer um passo muito curto e direto, mas não deixa de ser um passo.

Você fica no topo de um prédio alto, ao lado de um relógio de pêndulo com hora, minuto e ponteiro dos segundos. Em sua mão está uma bola de boliche e você a deixa cair do telhado. Em que tique-taque do relógio você ouvirá o estrondo da bola de boliche atingindo o chão?

Para responder com precisão, você deve usar crenças como A gravidade da Terra é de 9,8 metros por segundo e este edifício tem cerca de 120 metros de altura. Essas crenças não são antecipações sem palavras de uma experiência sensorial; elas são mais ou menos verbais, proposicionais. Provavelmente, não é exagero descrever essas duas crenças como sentenças feitas de palavras. Mas essas duas crenças têm uma consequência inferencial que é uma antecipação sensorial direta — se o ponteiro dos segundos do relógio estiver no numeral 12 quando você deixar cair a bola, você antecipa vê-lo no numeral 1 quando ouvir a batida cinco segundos depois. Para antecipar experiências sensoriais com a maior precisão possível, devemos processar crenças que não são antecipações de experiências sensoriais.

É uma grande força do Homo sapiens que possamos, melhor do que qualquer outra espécie no mundo, aprender a modelar o invisível. É também um dos nossos grandes pontos fracos. Os humanos geralmente acreditam em coisas que não são apenas invisíveis, mas irreais.

O mesmo cérebro que constrói uma rede de causas inferidas por trás da experiência sensorial também pode construir uma rede de causas não conectadas à experiência sensorial ou mal conectadas. Os alquimistas acreditavam que o flogisto causava fogo — poderíamos simplificar demais suas mentes desenhando um pequeno nodo denominado “flogisto” e uma flecha desse nodo para sua experiência sensorial de uma fogueira crepitante —, mas essa crença não fornecia previsões antecipadas; a ligação do flogisto à experiência sempre foi configurada após a experiência, em vez de restringir a experiência antecipadamente. Ou suponha que seu professor de inglês pós-moderno lhe ensine que o famoso escritor Wulky Wilkinsen é, na verdade, um “pós-utópico”. O que isso significa que você deve esperar de seus livros? Nada. A crença, se você pode

chamá-la assim, não se conecta à experiência sensorial de forma alguma. Mas é melhor você se lembrar da afirmação proposicional de que "Wulky Wilkinsen" tem o atributo "pós-utópico", então você pode regurgitar isso no próximo teste. Da mesma forma, se os "pós-utópicos" apresentam "alienação colonial"; se o questionário perguntar se Wulky Wilkinsen mostra alienação colonial, é melhor responder sim. As crenças estão conectadas umas às outras, embora ainda não estejam conectadas a nenhuma experiência antecipada.

Podemos construir redes inteiras de crenças conectadas apenas umas às outras — chamamos essas crenças de "flutuantes". É uma falha exclusivamente humana entre as espécies animais, uma perversão da capacidade do Homo sapiens de construir redes de crenças mais gerais e flexíveis.

A virtude racionalista do empirismo consiste em perguntar constantemente quais experiências nossas crenças preveem — ou melhor ainda, proíbem. Você acredita que o flogisto é a causa do fogo? Então, o que você espera ver acontecer, por causa disso? Você acredita que Wulky Wilkinsen é um pós-utópico? Então, o que você espera ver por causa disso? Não, não "alienação colonial"; que experiência vai acontecer com você? Você acredita que, se uma árvore cair na floresta e ninguém ouvir, ela ainda faz barulho? Então, que experiência deve acontecer com você?

Melhor ainda é perguntar: que experiência não deve acontecer com você? Você acredita que Elã vital explica a misteriosa vitalidade dos seres vivos? Então, o que essa crença não permite que aconteça — o que definitivamente falsificaria essa crença? Uma resposta nula significa que sua crença não restringe a experiência; permite que qualquer coisa aconteça com você. Flutua.

Ao discutir uma questão aparentemente factual, tenha sempre em mente sobre qual diferença de antecipação você está discutindo. Se você não consegue encontrar a diferença de antecipação, provavelmente está discutindo sobre rótulos em sua rede de crenças — ou pior ainda, crenças flutuantes, cracas em sua rede. Se você não sabe quais experiências estão implícitas em Wulky Wilkinsen ser um pós-utópico, você pode continuar discutindo para sempre.

Acima de tudo, não pergunte no que acreditar — pergunte no que antecipar. Toda questão de crença deve fluir de uma questão de antecipação, e essa questão de antecipação deve ser o centro da investigação. Cada suposição de crença deve começar fluindo para uma suposição específica de antecipação e deve continuar a pagar o aluguel em antecipações futuras. Se uma crença se tornar um caloteiro, descarte-a.

12 — Uma fábula de ciência e política

Durante o Império Romano, a vida cívica era dividida entre as facções Azul e Verde. Os Azuis e os Verdes se matavam em combates individuais, emboscadas, batalhas em grupo e motins. Procópio afirmou sobre as facções em guerra: "Cresce neles contra seus semelhantes uma hostilidade que não tem causa e em nenhum momento cessa ou desaparece, pois não dá lugar nem aos laços de casamento, nem de parentesco, nem de amizade, e o caso é o mesmo que aqueles que diferem com relação a essas cores sejam irmãos ou qualquer outro parente." [1] Edward Gibbon escreveu: "O apoio de uma facção tornou-se necessário para todo candidato a honras civis ou eclesiásticas." [2]

Mas quem eram os Azuis e os Verdes? Eles eram fãs de esportes — os partidários das equipes de corrida de bigas azuis e verdes.

Imagine uma sociedade futura que fugiu para uma vasta rede subterrânea de cavernas e selou suas entradas. Não especificaremos se eles fugiram de doenças, guerras ou radiação; suponhamos que os primeiros *Undergrounders* cultivaram alimentos, encontraram água, reciclaram ar, produziram luz e sobreviveram, e que seus descendentes prosperaram e eventualmente formaram cidades. Do mundo acima, só existem lendas escritas em pedaços de papel; e um desses pedaços descreve o céu, um vasto espaço aberto de ar acima de um grande chão ilimitado. O céu é de cor azul-celeste e contém estranhos objetos flutuantes, como enormes tufos de algodão branco. Mas o significado da palavra "cerúleo" é controverso; alguns dizem que se refere à cor conhecida como "azul" e outros que se referem à cor conhecida como "verde".

Nos primórdios da sociedade underground, Azuis e Verdes contestavam violentamente uns aos outros; mas hoje prevalece uma trégua — uma paz nascida de um crescente sentimento de inutilidade. Os costumes culturais mudaram; há uma grande e próspera classe média que cresceu com a aplicação eficaz da lei e não está acostumada à violência. As escolas fornecem alguma perspectiva histórica; quanto tempo durou a batalha entre Azuis e Verdes, quantos morreram, e quão pouco mudou como resultado. As mentes foram abertas para a estranha nova filosofia de que as pessoas são pessoas, sejam elas Azuis ou Verdes.

O conflito ainda não desapareceu. A sociedade permanece dividida nas linhas azul e verde, e há uma posição "azul" e uma "verde" em quase todas as questões contemporâneas de importância política ou cultural. Os Azuis defendem impostos sobre a renda individual, enquanto os Verdes defendem impostos sobre as vendas dos comerciantes; os Azuis defendem leis de casamento mais rígidas, enquanto os Verdes desejam facilitar a obtenção de divórcios; os Azuis obtêm seu apoio no coração das áreas urbanas, enquanto os fazendeiros e vendedores de água mais distantes tendem a ser Verdes; os Azuis acreditam que a Terra é uma enorme rocha esférica no centro do universo, enquanto os Verdes acreditam ser uma enorme rocha plana circulando algum outro objeto chamado Sol. Nem todo cidadão azul ou verde assume a posição "azul" ou "verde" em todas as questões, mas seria raro encontrar um comerciante da cidade que acreditasse que o céu é azul, e que, no entanto, defendesse um imposto individual e leis de casamento mais liberais.

O *Underground*[4] continua polarizado, uma paz inquieta. Algumas pessoas pensam genuinamente que Azuis e Verdes devem ser amigos, e agora é comum para um Verde patrocinar uma loja Azul, ou para um Azul visitar uma taverna Verde. No entanto, de uma trégua originalmente nascida da exaustão, há um espírito crescente de tolerância, até mesmo de amizade.

4 NT: Submundo. Mantido em inglês por se tratar de uma terra fictícia usada como alegoria pelo autor.

Certo dia, o *Underground* é abalado por um pequeno terremoto. Um grupo de seis pessoas é pego no tremor enquanto olha para as ruínas de antigas habitações nas cavernas superiores. Elas sentem o breve movimento da pedra sob seus pés, e uma das turistas tropeça e rala o joelho. A festa decide voltar, temendo novos terremotos. No caminho de volta, uma pessoa sente um cheiro estranho no ar, que vem de uma passagem há muito tempo não utilizada. Ignorando as advertências bem-intencionadas dos companheiros de viagem, a pessoa pega emprestada uma lanterna elétrica e entra na passagem. O corredor de pedra sobe… e sobe… e finalmente termina em um buraco escavado no mundo, um lugar onde toda pedra termina. A distância, uma distância infinita, se estende para sempre; um espaço de encontro para manter mil cidades. Inimaginavelmente alto, brilhando demais para olhar diretamente, uma faísca abrasadora lança luz sobre todo o espaço visível, o filamento nu de uma enorme lâmpada. No ar, pendurados sem suporte, estão grandes tufos incompreensíveis de algodão branco. E a vastidão brilhante acima… a cor é...

Agora a história se ramifica, dependendo de qual membro do grupo de turismo decidiu seguir o corredor até a superfície.

Aditya, do grupo Azul, permaneceu sob o céu azul para sempre e sorriu lentamente. Não era um sorriso amistoso. Havia ódio e orgulho ferido; lembrava todas as discussões que ela já teve com um Verde, todas as rivalidades, todas as promoções contestadas. “Você estava certa o tempo todo”, o céu sussurrou para ela, “e agora você pode provar isso”. Aditya ficou ali, absorvendo a mensagem, glorificando-se com ela, e então voltou para o corredor de pedra para contar ao mundo. Enquanto caminhava, ela fechou a mão em um punho. “A trégua acabou”, disse ela. Barron, do grupo Verde, olhou para o caos de cores sem compreender por longos segundos. Quando finalmente entendeu, sentiu um soco no estômago e lágrimas brotaram de seus olhos. Barron pensou no Massacre de Cathay, onde um exército Azul havia massacrado todos os cidadãos de uma cidade Verde, incluindo crianças; ele pensou no antigo general azul, Annas Rell, que havia declarado os Verdes “um poço de doença; uma pestilência a ser purificada”; ele pensou nos brilhos de ódio que viu nos olhos azuis e algo dentro dele se partiu. “Como você pode estar do lado deles?” Barron gritou para o céu e começou a chorar; porque ele sabia, sob o brilho azul malévolo, que o universo sempre foi um lugar maligno.

Charles, o Azul, observou o teto azul com surpresa. Como professor em uma faculdade mista, Charles enfatizou cuidadosamente que os pontos de vista Azul e Verde eram igualmente válidos e mereciam tolerância: o céu era uma construção metafísica e a cor cerúlea podia ser vista de mais de uma maneira. Charles se perguntou se, de onde um verde estivesse parado, ele não veria um teto verde acima, ou se talvez o teto fosse verde amanhã a esta hora, mas ele não podia apostar a sobrevivência da civilização nisso. Este era apenas um fenômeno natural, nada tendo a ver com filosofia, moral ou sociedade, mas uma que poderia ser facilmente mal interpretada, temia Charles. Suspirando, Charles se virou para voltar ao corredor. Amanhã, ele voltaria sozinho e bloquearia a passagem.

Daria, que antes era verde, tentou respirar em meio às cinzas de seu mundo. “Não vou recuar”, Daria disse a si mesma, “não vou desviar o olhar”. Ela foi verde toda a sua vida e agora ela precisava ser azul. Seus amigos e sua família se afastariam dela. “Fale a verdade, mesmo que sua voz trema”, seu pai lhe dissera. Mas seu pai estava morto agora e sua mãe nunca entenderia. Daria olhou para o céu azul, tentando aceitá-lo, e finalmente sua respiração se acalmou. “Eu estava errada”, ela disse tristemente para si mesma, “não é tão complicado, afinal”. Ela encontraria novos amigos e talvez sua família a perdoasse… ou ela se perguntou, com um tom de esperança, se poderia passar pelo mesmo teste de pé sob o mesmo céu. “O céu é azul”, disse Daria experimentalmente, e nada terrível aconteceu com ela, mas ela não conseguiu sorrir. Daria, a Azul, exalou tristemente e voltou ao mundo, imaginando o que diria.

Eddin, que era verde, olhou para o céu azul e começou a rir cinicamente. O curso da história de seu mundo finalmente ficou claro. Nem mesmo ele conseguia acreditar que eles tivessem sido tão tolos. “Estúpido”, disse Eddin, “estúpido, estúpido, e o tempo todo estava bem aqui”. Ódio, assassinatos, guerras e, o tempo todo, era apenas uma coisa em algum lugar, sobre a qual alguém havia escrito como escreveria sobre qualquer outra coisa. Nenhuma poesia, nenhuma beleza, nada com que qualquer pessoa sã se importasse, apenas uma coisa sem sentido que havia sido exagerada. Eddin encostou-se na boca da caverna, cansado, tentando pensar em uma maneira de evitar que essa informação explodisse o mundo, e se perguntando se eles não a mereciam.

Ferris se engasgou involuntariamente, paralisado de espanto e prazer. Seus olhos famintos dispararam em volta, fixando-se em cada visão antes de passar relutantemente para a próxima; o céu azul, as nuvens brancas, o vasto desconhecido lá fora, cheio de lugares, coisas (e pessoas?) que nenhum morador do subterrâneo jamais havia visto. “Ah, então é dessa cor”, disse Ferris, e partiu para explorar.

Referências

1. *Procopius, History of the Wars, ed. Henry B. Dewing, vol. 1 (Harvard University Press, 1914).*
2. *Edward Gibbon, The History of the Decline and Fall of the Roman Empire, vol. 4 (J. & J. Harper, 1829).*

13 — Crença na crença

Certa vez, Carl Sagan contou uma parábola sobre alguém que chega até nós e afirma: “Há um dragão na minha garagem”. Fascinante! Respondemos que queremos ver esse dragão — vamos imediatamente para a garagem! “Mas espere”, o interlocutor nos diz, “é um dragão invisível”.

Isso, como Sagan aponta, não torna a hipótese irrefutável. Talvez vamos até a garagem e, embora não vejamos nenhum dragão, ouvimos uma respiração pesada sem fonte visível; pegadas aparecem misteriosamente no chão; e os instrumentos mostram que algo na garagem está consumindo oxigênio e expirando dióxido de carbono.

Mas suponha que digamos ao interlocutor: “Ok, vamos até a garagem e ver se ouvimos uma respiração pesada”, e ele rapidamente responder que “não, este é um dragão inaudível. Então, propomos medir o dióxido de carbono no ar, e o interlocutor nos diz que o dragão não respira. Sugerimos jogar um saco de farinha no ar para detectar um dragão invisível, e o interlocutor imediatamente nos diz: “O dragão é permeável à farinha”.

Carl Sagan usou essa parábola para ilustrar a moral clássica de que hipóteses ruins precisam fazer muito trabalho (e, rápido) para evitar serem refutadas. No entanto, eu conto essa parábola para enfatizar um ponto diferente: o interlocutor deve ter um modelo preciso da situação em algum lugar de sua mente, porque pode antecipar, com antecedência, exatamente quais resultados experimentais precisarão ser justificados.

Alguns filósofos ficam muito confusos com esses cenários, perguntando: “O interlocutor realmente acredita haver um dragão presente ou não?” Como se o cérebro humano só tivesse espaço de armazenamento suficiente para representar uma crença de cada vez! A mente humana é mais complexa do que isso. Existem diferentes tipos de crenças; nem todas as crenças são antecipações diretas. O interlocutor claramente não espera ver nada incomum ao abrir a porta da garagem. Caso contrário, eles não dariam desculpas antecipadas. Também pode ser que o conjunto de crenças proposicionais do interlocutor contenha a crença “Há um dragão na minha garagem”. Para um racionalista, pode parecer que essas duas crenças devem colidir e entrar em conflito, mesmo que sejam de tipos diferentes. Porém, é um fato físico que você pode escrever “O céu é verde!” ao lado de uma foto de um céu azul sem que o papel exploda em chamas.

Supõe-se que a virtude racionalista do empirismo nos impeça de cometer esse tipo de erro. Devemos constantemente perguntar às nossas crenças quais experiências elas preveem e fazê-las pagar o aluguel antecipadamente. Mas o problema do interlocutor do dragão é mais profundo e não pode ser curado com um conselho tão simples. Não é exatamente difícil conectar a crença em um dragão com a experiência antecipada da garagem. Se você acredita que há um dragão em sua garagem, espera abrir a porta e ver um dragão. Se você não vê um dragão, isso significa não haver um dragão em sua garagem. É bastante simples e até mesmo pode ser testado em sua própria garagem.

Não, essa questão da invisibilidade é sintoma de algo muito pior.

Dependendo de como foi sua infância, você pode se lembrar de um momento em que começou a duvidar da existência do Papai Noel, mas ainda sentia que deveria acreditar nele, então tentou negar suas dúvidas. Como observa Daniel Dennett, quando é difícil acreditar em algo, muitas vezes é mais fácil acreditar que você deve acreditar. O que significa acreditar que o Ultimato Cósmico Celeste é perfeitamente azul e perfeitamente verde? Essa afirmação é confusa; não está claro o que exatamente seria acreditar, se você

acreditasse. No entanto, você pode acreditar com mais facilidade que é apropriado, benéfico, virtuoso e bom acreditar que o Ultimato Cósmico Celeste é perfeitamente azul e perfeitamente verde. Dennett chama essa forma de crença de "crença na crença". [1]

E aqui as coisas ficam complicadas, como as mentes humanas costumam fazer — acho até que Dennett simplifica demais como a psicologia funciona na prática. Por um lado, se você acredita em uma crença, não pode admitir para si mesmo que acredita apenas na crença, porque acreditar é virtuoso, não acreditar na crença, então, se você acredita apenas na crença, em vez de acreditar, você é não virtuoso. Ninguém admitirá para si mesmo: "Não acredito que o Céu Cósmico Supremo seja azul e verde, mas acredito que devo acreditar" — a menos que sejam extraordinariamente capazes de reconhecer sua própria falta de virtude. As pessoas não acreditam na crença, na crença, elas apenas acreditam na crença.

(Aqueles que acham isto confuso podem achar útil estudar a lógica matemática, que treina a pessoa a fazer distinções muito nítidas entre a proposição P, uma prova de P e uma prova de que P é demonstrável.) Existem distinções igualmente nítidas entre P, querer P, acreditar em P, querer acreditar em P e acreditar que você acredita em P.)

Existem diferentes tipos de crença na crença. Você pode acreditar explicitamente na crença; você pode recitar em sua mente consciente a frase verbal "É virtuoso acreditar que o Céu Cósmico Supremo é perfeitamente azul e perfeitamente verde" (enquanto também acredita que você acredita nisso, a menos que seja extraordinariamente capaz de reconhecer sua própria falta de virtude). Mas também existem formas menos explícitas de crença na crença. Talvez o defensor do dragão tema ser ridicularizado em público se confessar que estava errado (embora, na verdade, um racionalista o parabenize, e outros ridicularizem provavelmente o defensor se continuar afirmando haver um dragão em sua garagem). Pode ser que o defensor do dragão se esquive da perspectiva de admitir para si mesmo que não há nenhum dragão, porque entra em conflito com sua autoimagem de glorioso descobridor do dragão, que viu em sua garagem o que todos os outros não conseguiram ver.

Se todos os nossos pensamentos fossem sentenças verbais deliberadas, como os filósofos manipulam, a mente humana seria muito mais fácil de entender para os humanos. Imagens mentais fugazes, hesitações não expressas, desejos ativados sem reconhecimento — tudo isso faz parte de nós tanto quanto as palavras.

Embora eu discorde de Dennett em alguns detalhes e complicações, ainda acho que a noção de crença na crença é o insight chave, necessário para entender o interlocutor do dragão. No entanto, precisamos de um conceito mais amplo de crença, que não se limite apenas a frases verbais. A palavra "crença" deve incluir também controladores de antecipação tácitos, e a "crença na crença" deve incluir guias cognitivo-comportamentais tácitos. Não é psicologicamente realista afirmar que "O interlocutor do dragão não acredita que haja um dragão em sua garagem; eles acreditam ser benéfico acreditar que há um dragão em sua garagem". No entanto, é realista afirmar que o interlocutor do dragão antecipa como se não houvesse dragão em sua garagem e dá desculpas como se acreditasse na crença.

Você pode ter uma imagem mental comum de sua garagem, sem dragões nela, que prevê corretamente suas experiências ao abrir a porta e nunca pensar na frase verbal Não há dragão em minha garagem. Eu até aposto que já aconteceu com você — que quando você abre a porta da garagem ou do quarto, ou qualquer outro lugar, e espera não ver dragões, nenhuma frase verbal passa pela sua mente.

Para manter a crença no dragão, ou para preservar sua autoimagem como alguém que acredita no dragão, não é necessário pensar explicitamente que você quer acreditar que há um dragão em sua garagem. Basta recuar diante da perspectiva de admitir que não acredita.

Para prever corretamente, com antecedência, quais resultados experimentais devem ser justificados, o requerente do dragão deve (a) possuir um modelo de controle de antecipação preciso em algum lugar de sua mente e (b) agir cognitivamente para proteger (b1) sua liberdade — crença proposicional flutuante no dragão ou (b2) sua autoimagem de acreditar no dragão.

Se alguém acredita em sua crença no dragão e também acredita no dragão, o problema é muito menos grave. Eles estarão dispostos a arriscar o pescoço em previsões experimentais e talvez até concordem em desistir da crença se a previsão experimental estiver errada — embora a crença na crença ainda possa interferir nisso, se a crença em si não for absolutamente confiável. Quando alguém inventa desculpas antecipadamente, parece exigir que a crença e a crença na crença tenham se tornado dessincronizadas.

Referências

1. *Daniel C. Dennett, Breaking the Spell: Religion as a Natural Phenomenon (Penguin, 2006).*

14 — Judô bayesiano

É possível se divertir com pessoas cujas expectativas estão em desacordo com o que acreditam acreditar.

Certa vez, estava em um jantar e tentava explicar minha profissão a um homem, quando ele disse: “Não acredito que a Inteligência Artificial seja possível, porque só Deus pode criar uma alma”.

Naquele momento, algo divino deve ter me inspirado, pois respondi instantaneamente: “Quer dizer que, se eu puder criar uma Inteligência Artificial, isso provará que sua religião é falsa?”

Ele disse: “O quê?”

Eu disse: “Se sua religião prevê que eu não possa criar uma Inteligência Artificial, então, se eu criar uma, isso significa que sua religião é falsa. Ou sua religião permite que eu crie uma IA, ou, se eu criar uma IA, isso refutará sua religião.”

Houve uma pausa, quando ele percebeu que acabara de tornar sua hipótese vulnerável à refutação , e então disse: “Bem, eu não quis dizer que você não poderia criar uma inteligência, apenas que não poderia ser emocional da mesma forma que nós”.

Eu disse: “Então, se eu criar uma Inteligência Artificial que, sem ser deliberadamente pré-programada com qualquer tipo de roteiro, comece a falar sobre uma vida emocional que soa como a nossa, isso significa que sua religião está errada”.

Ele disse: “Bem, acho que teremos que concordar em discordar sobre isso”.

Eu disse: “Não, na verdade, não podemos. Existe um teorema da racionalidade chamado Teorema do Acordo de Aumann, que mostra que dois racionalistas não podem concordar em discordar. Se duas pessoas discordam uma da outra, pelo menos uma delas deve estar cometendo um erro”.

Conversamos rapidamente sobre isso. Finalmente, ele disse: “Bem, acho que estava realmente tentando dizer que não acredito que você possa tornar algo eterno”.

Eu disse: “Bem, eu também não acredito! Estou feliz que pudemos concordar sobre isso, como exige o Teorema de Acordo de Aumann.” Estendi minha mão, ele a apertou e depois se afastou.

Uma mulher que estava por perto e ouvindo a conversa, disse-me gravemente: “Isso foi lindo”.

“Muito obrigado,” eu disse.

15 — Fingindo ser sábio

"O lugar mais quente do Inferno é reservado para aqueles que, em tempos de crise, permanecem neutros[5]."

—~~Dante Alighieri, famoso especialista em infernos~~ John F. Kennedy, aquele que cita erroneamente

É comum mostrar neutralidade ou julgamento suspenso para indicar que alguém é maduro, sábio, imparcial ou tem um ponto de vista superior. Por exemplo, meus pais respondem a perguntas teológicas como "Por que o antigo Egito, que tinha bons registros em muitos outros assuntos, carece de registros de judeus que já estiveram lá?" com "Oh, quando eu tinha a sua idade, eu também costumava fazer esse tipo de pergunta, mas agora eu superei isso."

Outro exemplo seria o da diretora que, diante de duas crianças flagradas brigando no parquinho, diz com severidade: "Não importa quem começou a briga, importa apenas quem a termina." Claro que importa quem começou a briga. A diretora pode não ter acesso a boas informações sobre esse fato crítico, mas se tiver, a diretora deve dizê-lo, sem descartar a importância de quem deu o primeiro soco. Deixe um pai tentar dar um soco no diretor e veremos até onde "não importa quem começou" convence o juiz. Mas para os adultos é apenas inconveniente que as crianças briguem, e não importa para sua conveniência qual criança começou. É apenas conveniente que a luta termine o mais rápido possível.

Uma dinâmica semelhante, acredito, rege as situações na diplomacia internacional nas quais as grandes potências exigem severamente que grupos menores cessem as hostilidades imediatamente. Não importa para a Grande Potência quem começou, quem provocou ou quem respondeu desproporcionalmente à provocação, porque a inconveniência contínua da Grande Potência é apenas uma função do conflito em andamento. Oh, Israel e o Hamas, por que vocês não podem simplesmente se dar bem?

Isso que chamo de "fingir ser sábio". Claro, existem muitas maneiras de tentar mostrar sabedoria. Mas tentar mostrar sabedoria recusando-se a supor, resumir evidências, julgar ou tomar partido, ficando acima da briga e olhando com um olhar superior e condescendente — ou seja, mostrando sabedoria apenas com palavras e não fazendo nada — isso, eu acho particularmente pretensioso.

Paulo Freire disse: "Lavar as mãos do conflito entre os poderosos e os impotentes significa ficar do lado dos poderosos, não ser neutro." Um parquinho é um ótimo lugar para ser um valentão e um lugar terrível para ser uma vítima, se os professores não se importam com quem começou. E da mesma forma na política internacional: um mundo onde as grandes potências se recusam a tomar partido e apenas exigem tréguas imediatas é um grande mundo para os agressores e um lugar terrível para os agredidos. Mas, claro, é um mundo muito conveniente para ser uma Grande Potência ou um diretor de escola.

Portanto, parte desse comportamento pode ser atribuída ao puro egoísmo por parte dos "sábios".

Mas parte disso também tem a ver com mostrar um ponto de vista superior. Afinal, o que os outros adultos pensariam de uma diretora que parecesse estar tomando partido em uma briga entre crianças? Isso diminuiria o status da diretora a uma mera participante da briga!

---

5 NT: Tradução livre do texto original em inglês. *The hottest place in Hell is reserved for those who in time of crisis remain neutral.*

Da mesma forma, com o ancião respeitado — que pode ser um CEO, um acadêmico de prestígio ou o fundador de uma lista de mala direta — cuja reputação de imparcialidade depende da recusa em julgar a si mesmos quando outros estão escolhendo um lado.

Os lados buscam o apoio dos sábios, mas quase sempre em vão, pois eles são juízes respeitados com a condição de que raramente realmente julgam — assim, eles seriam apenas mais um disputante na briga, não melhores do que qualquer outro argumentador.

(Estranhamente, os juízes no sistema jurídico atual podem proferir repetidos vereditos sem perder automaticamente sua reputação de imparcialidade.) Talvez seja devido à norma de que eles têm que julgar, o qual é o trabalho deles. Ou talvez seja porque os juízes não precisam decidir repetidamente sobre questões que dividem uma tribo da qual dependem para sua reverência.)

Há ocasiões em que é racional suspender o julgamento, em que as pessoas julgam rapidamente devido a seus preconceitos. Como Michael Rooney disse:

O erro aqui é semelhante a um que vejo frequentemente em estudantes iniciantes de filosofia: quando confrontados com motivos para serem céticos, eles se tornam relativistas. Ou seja, quando a conclusão racional é suspender o julgamento sobre uma questão, muitas pessoas concluem que qualquer julgamento é tão plausível quanto qualquer outro.[6]

Mas como podemos evitar o comportamento pseudo-racionalista (relacionado, mas distinto) de sinalizar sua imparcialidade imparcial alegando falsamente que o atual equilíbrio de evidências é neutro? “Ah, bem, é claro que existem muitos darwinistas apaixonados por aí, mas acredito que as evidências que temos realmente não nos permitem fazer um endosso definitivo da seleção natural sobre o design inteligente”.

Nesse ponto, recomendo lembrar que a neutralidade é um julgamento definitivo. Não significa que não se tenha opinião sobre nada. Significa apresentar a posição definida e particular de que o equilíbrio das evidências em um caso particular permite apenas uma soma, que é neutra. Isso também pode estar errado; propor a neutralidade é tão atacável quanto propor qualquer lado em particular.

Da mesma forma, em questões políticas. Se alguém diz que os lados pró-vida e pró-escolha têm pontos positivos, que deveriam tentar comprometer-se e respeitar mais um ao outro, não estão assumindo uma posição acima dos dois lados padrão no debate sobre o aborto. Estão apresentando um julgamento definitivo, tão particular quanto dizer “pró-vida!” ou “pró-escolha!”.

Se seu objetivo é melhorar sua capacidade geral de formar crenças mais precisas, pode ser útil evitar questões emocionalmente carregadas, como aborto ou conflito israelense-palestino. Mas isso não significa que um racionalista seja maduro demais para falar de política. Não significa que um racionalista esteja acima dessa briga tola na qual apenas meros partidários políticos e jovens entusiastas se rebaixariam a participar.

Conforme Robin Hanson, a habilidade de ter conversas potencialmente divisivas é um recurso limitado. Se você puder pensar em maneiras de puxar a corda para o lado, é justificável gastar seus recursos limitados em questões menos comuns, onde a discussão marginal oferece recompensas marginais relativamente maiores.

No entanto, as responsabilidades que você desprioriza são uma questão de seus recursos limitados, não uma questão de flutuar acima de tudo, sereno e sábio.

---

6 NT: Tradução livre do texto original. *The error here is similar to one I see all the time in beginning philosophy students: when confronted with reasons to be skeptics, they instead become relativists. That is, when the rational conclusion is to suspend judgment about an issue, all too many people instead conclude that any judgment is as plausible as any other.*

Minha resposta ao comentário de Paul Graham no Hacker News é um resumo que vale a pena repetir:

Existe uma diferença entre:

- Fazer um julgamento neutro;
- Recusar-se a investir recursos marginais;
- Fingir que qualquer um dos itens acima é uma marca de profunda sabedoria, maturidade e um ponto de vista superior; com a implicação correspondente de que os lados originais ocupam pontos de vista inferiores que não são muito diferentes do seu.

Referências

1. *Paulo Freire, The Politics of Education: Culture, Power, and Liberation (Greenwood Publishing Group, 1985), 122.*

16 — A afirmação da religião de ser não refutável

A narrativa mais antiga que conheço de um experimento científico é, ironicamente, a história de Elias e os sacerdotes de Baal.

O povo de Israel estava indeciso entre Jeová e Baal, então Elias propôs um experimento para resolver a questão — um conceito bastante novo naquela época! Os sacerdotes de Baal colocariam seu touro em um altar e Elias colocaria o touro de Jeová em outro altar, mas nenhum dos lados poderia acender o fogo; qualquer Deus que fosse real invocaria fogo sobre seu sacrifício. Os sacerdotes de Baal serviam como grupo de controle para Elias — o mesmo combustível de madeira, o mesmo touro e os mesmos sacerdotes fazendo invocações, mas a um deus falso. Em seguida, Elias derramou água em seu altar, comprometendo a simetria experimental, mas isso foi nos primeiros dias — para significar a aceitação deliberada do ônus da prova, como a necessidade de um nível de significância de 0,05. O fogo desceu sobre o altar de Elias, a qual foi a observação experimental. O povo de Israel que assistia gritou "O Senhor é Deus!" —revisão por pares.

Em seguida, o povo arrastou os 450 sacerdotes de Baal até o rio Quisom e cortou suas gargantas. Isso é severo, mas necessário. É preciso descartar firmemente a hipótese refutada e fazer isso rapidamente, antes que ela possa gerar desculpas para se proteger. Se os sacerdotes de Baal sobrevivessem, começariam a tagarelar sobre como a religião é um magistério separado que não pode ser provado nem refutado.

Antigamente, as pessoas realmente acreditavam em suas religiões, em vez de apenas afirmarem acreditar. Os arqueólogos bíblicos que procuraram a Arca de Noé não pensaram estarem perdendo tempo; eles previram que poderiam se tornar famosos. Somente depois de não encontrarem evidências confirmatórias — e, ao contrário disso, encontrarem evidências refutadas—é que os religiosos adotaram o que William Bartley chamou de retirada para o compromisso: "Eu acredito porque acredito".

Nos velhos tempos, não havia o conceito de que a religião era um magistério separado. O Antigo Testamento é um depósito de cultura de fluxo de consciência: história, leis, parábolas morais e, sim, modelos de como o universo funciona. Em nenhuma passagem do Antigo Testamento, você encontrará alguém falando sobre uma maravilha transcendente na complexidade do universo. Mas você encontrará muitas afirmações científicas, como o universo criado em seis dias (uma metáfora para o Big Bang) ou coelhos ruminando (uma metáfora para…).

Antigamente, dizer que a religião local "não podia ser provada" poderia resultar em ser queimado na fogueira. Uma das crenças centrais do judaísmo ortodoxo é que Deus apareceu no Monte Sinai e disse com uma voz trovejante: "Sim, é tudo verdade". De uma perspectiva bayesiana, essa é uma evidência inequívoca de uma entidade poderosa sobre-humana. (Embora não prove que a entidade é Deus, por assim dizer, ou que a entidade é benevolente — podem ser adolescentes alienígenas ao invés de Deus.) A grande maioria das religiões na história humana — exceto aquelas inventadas muito recentemente — conta histórias de eventos que seriam evidências inequívocas se eles tivessem realmente acontecido.

A ortogonalidade entre religião e fatos é um conceito recente e estritamente ocidental. As pessoas que escreveram as escrituras originais não sabiam dessa diferença. O Império Romano herdou a filosofia dos antigos gregos, impôs a lei e a ordem em suas províncias, manteve registros burocráticos e tolerou a liberdade religiosa. O Novo Testamento, criado durante a época do Império Romano, traz consigo alguns traços de modernidade. Você não poderia inventar uma história sobre Deus destruindo completamente a cidade de Roma (como em Sodoma e Gomorra), porque os historiadores romanos questionariam sua história, e você não poderia simplesmente apedrejá-los.

Em contraste, as pessoas que inventaram as histórias do Antigo Testamento podiam inventar praticamente qualquer coisa que quisessem. Os primeiros egiptólogos ficaram genuinamente chocados ao não encontrar nenhum vestígio de tribos hebraicas que tivessem estado no Egito — eles não esperavam encontrar um registro das Dez Pragas, mas esperavam encontrar algo. Como se viu, eles encontraram algo. Eles descobriram que, durante a suposta época do êxodo, o Egito governava grande parte de Canaã. Isso é um grande erro histórico, mas se não houver bibliotecas, ninguém poderá questioná-lo.

O Império Romano tinha bibliotecas. Portanto, o Novo Testamento não reivindica milagres geopolíticos grandes, vistosos e de larga escala, como o Antigo Testamento costumava fazer. Em vez disso, o Novo Testamento reivindica milagres menores que, no entanto, se encaixam na mesma estrutura de evidência. Um menino cai e começa a espumar pela boca; a causa é um espírito imundo; seria razoável esperar que um espírito imundo fugisse de um profeta verdadeiro, mas não de um charlatão; Jesus expulsa o espírito imundo; portanto, Jesus é um profeta verdadeiro e não um charlatão. Esse raciocínio segue a lógica bayesiana, mas pressupõe a crença de que a epilepsia é causada por demônios e que o fim do ataque é prova de que o demônio foi expulso.

A religião não só costumava fazer afirmações sobre questões factuais e científicas, como também costumava fazer afirmações sobre tudo. A religião estabeleceu um código de leis, antes dos corpos legislativos; a religião estabeleceu a história — antes de historiadores e arqueólogos; a religião estabeleceu a moral sexual antes do movimento de libertação das mulheres; a religião descrevia as formas de governo — antes das constituições; e a religião respondia a questões científicas desde a taxonomia biológica até a formação das estrelas. O Antigo Testamento não fala sobre um sentimento de admiração pela complexidade do universo, ele estava ocupado estabelecendo a pena de morte para mulheres que usavam roupas masculinas, o que era um conteúdo religioso sólido e satisfatório para aquela época. O conceito moderno de religião como puramente ética deriva de todas as outras áreas terem sido assumidas por instituições melhores. Ética é o que resta.

Ou melhor, as pessoas acham que a ética é o que resta. Imagine um depósito de cultura de 2.500 anos atrás. Com o tempo, a humanidade progredirá imensamente e as peças do antigo depósito de cultura se tornarão cada vez mais obsoletas. A ética não tem sido imune ao progresso humano — por exemplo, agora desaprovamos práticas aprovadas pela Bíblia, como manter escravos. Por que as pessoas pensam que a ética ainda é um jogo justo?

Intrinsecamente, não há nada de pequeno no problema ético de massacrar milhares de inocentes primogênitos do sexo masculino, para convencer um faraó não eleito a libertar escravos que logicamente poderiam ter sido teletransportados para fora do país. Deveria ser mais flagrante do que o erro científico, comparativamente trivial, de dizer que os gafanhotos têm quatro patas. No entanto, se você disser que a Terra é plana, as pessoas vão olhar para você como se você fosse louco. Mas se você disser que a Bíblia é sua fonte de ética, as mulheres não vão esbofeteá-lo. O conceito de racionalidade da maioria das pessoas é determinado pelo que elas acham que podem fazer; elas acham que podem se safar endossando a ética bíblica; e, portanto, isso requer apenas um esforço administrável de autoengano para elas ignorarem os problemas morais da Bíblia. Todos concordaram em não notar o elefante branco na sala de estar, e esse estado de coisas pode se sustentar por um tempo.

Talvez um dia a humanidade avance ainda mais, e qualquer pessoa que endosse a Bíblia como fonte de ética será tratada da mesma forma que Trent Lott endossou a campanha presidencial de Strom Thurmond. E, então, será dito que o "verdadeiro núcleo" da religião sempre foi a genealogia ou algo parecido.

A ideia de que a religião é um magistério separado que não pode ser provado ou refutado é uma grande mentira — uma mentira, continuamente repetida para que as pessoas a repitam sem pensar, mesmo que, sob um exame crítico, ela seja simplesmente falsa. Isso distorce gravemente como a religião se desenvolveu historicamente, como todas as escrituras apresentam suas crenças, o que as crianças ouvem para serem persuadidas e o que a maioria das pessoas religiosas ainda acredita. É admirável a ousadia daqueles que defendem essa mentira, semelhante à afirmação de que a Oceania sempre esteve em guerra com o Leste Asiático. O promotor saca o machado ensanguentado, e o réu, momentaneamente chocado, pensa rapidamente e diz: "Mas você não pode refutar minha inocência com meras evidências — é um magistério separado!"

E se isso não funcionar, pegue um pedaço de papel e rabisque um passe livre para você mesmo.

17 — Professar e torcer

Uma vez participei de um painel que debatia se ciência e religião são compatíveis. Uma das mulheres no painel, uma pagã, discorreu interminavelmente sobre como a Terra havia sido criada. Ela acreditava que a Terra foi criada quando uma vaca gigante primordial surgiu no abismo primordial e lambeu um deus primordial. Seus descendentes mataram um gigante primordial e usaram seu corpo para criar a Terra, entre outras coisas. A história era extensa, detalhada e mais absurda do que a Terra apoiada nas costas de uma tartaruga-gigante. E o orador claramente conhecia ciência o suficiente para saber disso.

Ainda me pego lutando para encontrar palavras para descrever o que vi quando essa mulher falou. Ela falou com… orgulho? Autossatisfação? Uma ostentação deliberada de si mesma?

A mulher continuou descrevendo seu mito da criação, pelo que pareceu uma eternidade, mas provavelmente durou apenas cinco minutos. Aquela estranha sensação de orgulho/satisfação/ostentação claramente tinha algo a ver com ela saber que suas crenças eram cientificamente ultrajantes. E não é que ela odiasse ciência; como palestrante, ela afirmou que religião e ciência eram compatíveis. Ela até mencionou que era bastante compreensível que os vikings falassem sobre um abismo primordial, dada a terra em que viviam — explicando sua própria religião! — e ainda assim insistiu que era nisso que ela "acreditava", com uma satisfação peculiar.

Não tenho certeza de que o conceito de "crença na crença" de Daniel Dennett se estenda para cobrir este evento. Era mais estranho do que isso. Ela não recitou seu mito da criação com a fé fanática de alguém que precisa se tranquilizar. Ela não agiu como se esperasse que nós, o público, fôssemos convencidos — ou como se precisasse de nossa crença para validá-la.

Dennett, além de sugerir a crença na crença, também sugeriu que muito do que é chamado de "crença religiosa" deveria ser estudado como "profissão religiosa". Suponha que um antropólogo alienígena estudasse um grupo de estudantes ingleses pós-modernos que aparentemente acreditavam que Wulky Wilkensen era um autor pós-utópico. A pergunta apropriada pode não ser "Por que todos os alunos acreditam nessa estranha crença?", mas "Por que todos eles escrevem essa frase estranha em questionários?" Mesmo que uma frase seja essencialmente sem sentido, você ainda pode saber quando deve entoar a resposta em voz alta.

Acho que Dennett pode ser um pouco cínico demais ao sugerir que professar uma crença religiosa é apenas a dizerem voz alta — a maioria das pessoas é honesta o suficiente para que, se fizerem uma declaração religiosa em voz alta, também se sentirão obrigadas a repetir essa mesma declaração em seu próprio fluxo de consciência.

Mas mesmo o conceito de "professar uma crença religiosa" não parece abranger a afirmação da mulher pagã sobre acreditar na vaca primordial. Se você tivesse que professar uma crença religiosa para satisfazer um padre, ou satisfazer um correligionário — diabos, para satisfazer sua própria autoimagem como uma pessoa religiosa — você teria que fingir acreditar de forma muito mais convincente do que esta mulher estava fazendo. Enquanto ela recitava a história da vaca primordial, com aquele mesmo orgulho ostentoso, ela nem tentava ser persuasiva, nem tentava nos convencer de que levava sua própria religião a sério. Acho que essa é a parte que me surpreendeu. Conheço pessoas que acreditam em coisas ridículas, mas quando as professam, esforçam-se muito mais para se convencer de que levam suas crenças a sério.

Finalmente, ocorreu-me que essa mulher não estava tentando nos convencer, ou mesmo convencer a si mesma. Sua recitação da história da criação não era sobre a criação do mundo. Em vez disso, ao lançar uma diatribe de cinco minutos sobre a vaca primordial, ela estava torcendo pelo paganismo, como se erguesse uma bandeira em um jogo de futebol. Um banner dizendo “Vamos lá, Azuis!” não é uma declaração de fato ou uma tentativa de persuadir; não precisa ser convincente — é uma expressão de alegria.

O estranho orgulho ostentoso que ela exibia era como se estivesse marchando nua em uma parada do orgulho gay. (Não que haja algo de errado em marchar nua em uma parada do orgulho gay.) A orientação sexual não é nada que a verdade possa destruir.) Não era apenas uma celebração, como marchar, mas uma celebração escandalosa, como marchar nua — acreditando que não poderia ser presa ou criticada, porque estava fazendo isso em nome do orgulho gay.

Por isso, era importante para ela que o que estava dizendo fosse absurdo. Se ela tentasse torná-lo mais plausível, seria como marchar vestida.

18 — Crença como vestimenta

Até agora, eu distingui entre crença como controladora de expectativa, crença na crença, professar e torcer. Dentre elas, podemos chamar as crenças que controlam a expectativa de "crenças apropriadas" e as outras formas de "crença inadequada". Uma crença apropriada pode estar errada ou ser irracional, como quando alguém realmente espera que a oração cure seu bebê doente. Mas as outras formas, indiscutivelmente, "não são crenças".

Outra forma de crença inadequada é a crença como identificação de grupo — como uma forma de pertencimento. Robin Hanson usa a excelente metáfora de usar roupas incomuns, um uniforme de grupo como as vestes de um padre ou um quipá judaico, e por isso chamarei isso de "crença como vestimenta".

Em termos de psicologia realista humana, os muçulmanos que pilotaram aviões contra o World Trade Center, sem dúvida, se viam como heróis defendendo a verdade, a justiça e o caminho islâmico de horrendos monstros alienígenas, como no filme *Independence Day*. Apenas um nerd muito inexperiente, do tipo que não tem ideia de como os não nerds veem o mundo, diria isso em voz alta em um bar no Alabama. Não é uma coisa americana de se dizer. A coisa americana de se dizer é que os terroristas "odeiam nossa liberdade" e que lançar um avião contra um prédio é um "ato covarde". Você não pode dizer as palavras "autossacrifício heroico" e "homem-bomba" na mesma frase, mesmo para descrever com precisão como o Inimigo vê o mundo. O próprio conceito da coragem e altruísmo de um homem-bomba é a vestimenta inimiga — dá para perceber, porque o inimigo fala sobre isso. A covardia e a sociopatia de um homem-bomba são a vestimenta americana. Não há aspas que você possa usar para falar sobre como o Inimigo vê o mundo; seria como se fantasiar de nazista para o Halloween.

A crença como vestimenta pode ajudar a explicar como as pessoas podem ser apaixonadas por crenças inadequadas. A mera crença na crença, ou a profissão religiosa, teria alguns problemas para criar efeitos emocionais genuínos, profundos e poderosos. Pelo menos é o que eu suponho. Confesso que não sou um especialista no assunto. Mas a minha impressão é esta: as pessoas que pararam de ter expectativas como se sua religião fosse verdadeira farão de tudo para se convencer de que são apaixonadas, e esse desespero pode ser confundido com paixão. Porém, esse não é o mesmo fervor que eles tinham quando crianças.

Por outro lado, é muito fácil para um ser humano pertencer a um grupo genuinamente, apaixonada e profundamente, torcer por seu time favorito. (Esta é a base sobre a qual repousa a fraude de "Republicanos x Democratas" e falsos dilemas análogos em outros países, mas isso é assunto para outra ocasião.) A identificação com uma tribo é uma força emocional muito forte. As pessoas estão dispostas a morrer por isso. E uma vez que você faz as pessoas se identificarem com uma tribo, as crenças que são a vestimenta dessa tribo serão expressas com toda a paixão de pertencer a essa tribo.

19 — Luzes de aplauso

No *Singularity Summit 2007*, um dos palestrantes pediu pelo desenvolvimento democrático e multinacional da Inteligência Artificial. Aproximei-me do microfone e perguntei:

"Suponha que um grupo de repúblicas democráticas forme um consórcio para desenvolver a IA, e haja muita politicagem durante o processo — alguns grupos de interesse têm uma influência extraordinariamente grande, outros são prejudicados — em outras palavras, o resultado se parece com os produtos das democracias modernas. Como alternativa, suponha que um grupo de nerds rebeldes desenvolva uma IA em seu porão e instrua a IA a fazer uma pesquisa para consultar a opinião de todas as pessoas no mundo — entregando celulares a qualquer um que não os tenha para poderem participar dessa pesquisa — e, de posse dos resultados, faça o que a maioria disser. Qual desses você acha que é mais "democrático" e com qual se sentiria mais seguro?"

Eu queria descobrir se ele acreditava na adequação pragmática do processo político democrático ou se acreditava na retidão moral do voto. Mas o palestrante respondeu:

"O primeiro cenário parece um editorial da revista *Reason*, e o segundo parece o enredo de um filme de Hollywood."

Confuso, perguntei:

"Então, que tipo de processo democrático você tinha em mente?"

O palestrante respondeu:

"Algo como o Projeto Genoma Humano — que era um projeto de pesquisa patrocinado internacionalmente."

Perguntei:

"Como diferentes grupos de interesse resolveriam seus conflitos em uma estrutura como o Projeto Genoma Humano?"

E o palestrante disse:

"Não sei."

Essa troca me faz lembrar de uma citação de algum ditador, a quem foi perguntado se ele tinha alguma intenção de levar seu estado de estimação para a democracia:

"Acreditamos que já estamos em um sistema democrático. Ainda faltam alguns fatores, como a expressão da vontade do povo."

A essência de uma democracia é o mecanismo específico que resolve conflitos políticos. Se todos os grupos tivessem as mesmas políticas preferidas, não haveria necessidade de democracia — cooperaríamos automaticamente. O processo de resolução pode ser um voto de maioria direta, ou uma legislatura eleita, ou mesmo um comportamento sensível ao eleitor de uma Inteligência Artificial, mas tem que ser algo. O que significa pedir uma solução "democrática" se você não tem em mente um mecanismo de resolução de conflitos?

Acredito que isso signifique que você mencionou a palavra “democracia” e, portanto, o público deveria aplaudir. Não é tanto uma declaração proposicional, mas o equivalente a uma luz de “Aplausos” que indica ao público do estúdio quando deve aplaudir.

Este caso é notável apenas porque eu confundi a luz dos aplausos com uma sugestão política, causando embaraço para todos. A maioria das luzes de aplausos é muito mais evidente e pode ser detectada facilmente. Por exemplo, se alguém disser:

Precisamos equilibrar os riscos e oportunidades da IA. Se você inverter essa afirmação, obterá:

Não devemos equilibrar os riscos e oportunidades da IA.

Como a inversão parece anormal, a declaração original provavelmente é normal, o que implica que não transmite novas informações. Existem muitas razões legítimas para fazer uma declaração que, isoladamente, não seria informativa. “Precisamos equilibrar os riscos e oportunidades da IA” pode introduzir um tópico de discussão, pode enfatizar a importância de uma proposta específica de equilíbrio ou pode criticar uma proposta desequilibrada. A vinculação a uma declaração normal pode transmitir novas informações para um racionalista limitado — a vinculação em si pode não ser óbvia. Mas se não houver detalhes a seguir, a frase é provavelmente apenas uma luz de aplausos.

Às vezes, sinto vontade de fazer uma palestra que consista apenas em aplausos e ver quanto tempo leva para o público começar a rir:

Estou aqui para propor a vocês hoje que seja necessário equilibrar os riscos e as oportunidades da Inteligência Artificial avançada. Devemos evitar os perigos desnecessários e aproveitar ao máximo as oportunidades que surgem. Para alcançar esses objetivos, devemos planejar com sabedoria e racionalidade, sem agir com medo e pânico ou ceder à tecnofobia. Também não devemos agir com entusiasmo cego. É essencial que respeitemos os interesses de todas as partes interessadas na Singularidade, buscando garantir que os benefícios das tecnologias avançadas sejam acessíveis ao maior número possível de pessoas, em vez de restringi-los a poucos. Devemos tentar evitar conflitos violentos que possam surgir do uso dessas tecnologias, e impedir que uma capacidade destrutiva massiva caia nas mãos erradas. É crucial refletirmos sobre essas questões antes que seja tarde demais para agir.

C — Percebendo Confusão

20 — Concentre sua incerteza

Os rendimentos dos títulos aumentarão, diminuirão ou permanecerão iguais? Se você é um comentarista de TV cujo trabalho é explicar o resultado depois que ele ocorre, não há motivo para se preocupar. Não importa qual das três possibilidades se torne realidade, você poderá explicar por que o resultado se encaixa perfeitamente em sua teoria do mercado de animais de estimação. Não há razão para pensar nessas três possibilidades como sendo de alguma forma opostas ou exclusivas uma da outra, porque você receberá uma nota alta como especialista, independentemente do resultado.

Mas espere! Suponha que você seja um comentarista de TV iniciante e não tenha experiência suficiente para inventar explicações plausíveis na hora. Você precisa preparar comentários com antecedência para a transmissão de amanhã e tem tempo limitado para se preparar. Nesse caso, seria útil saber qual resultado realmente ocorrerá — se os rendimentos dos títulos aumentarão, diminuirão ou permanecerão iguais — porque então você só precisaria preparar um conjunto de desculpas.

Infelizmente, ninguém pode prever o futuro. O que você pode fazer? Certamente, não pode usar "probabilidades". Todos aprendemos na escola que "probabilidades" são pequenos números que aparecem ao lado de um problema de palavras, e não há números pequenos aqui. Pior ainda, você se sente inseguro. Você não se lembra de ter se sentido inseguro ao manipular pequenos números em problemas de palavras.

As aulas da faculdade que ensinam matemática são lugares limpos e agradáveis, portanto, a matemática em si não pode ser aplicada às situações da vida que não são limpas e agradáveis. Você não gostaria de transferir habilidades de pensamento inadequadamente de um contexto para outro. Claramente, não se trata de "probabilidades".

No entanto, você tem apenas 100 minutos para preparar suas desculpas. Você não pode gastar todos os 100 minutos em "para cima", todos os 100 minutos em "para baixo" e todos os 100 minutos em "igual". Você precisa priorizar de alguma forma.

Se você precisasse justificar seu uso de tempo a um comitê de revisão, teria que gastar o mesmo tempo em cada possibilidade. Como não há números pequenos anotados, você não teria documentação para justificar o uso de diferentes quantidades de tempo. Você pode ouvir os revisores agora: Por que, Sr. Finkledinger, você gastou exatamente 42 minutos na desculpa número 3? Por que não 41 minutos ou 43? Admita — você não está sendo objetivo! Você está jogando com favoritismos subjetivos!

Mas, com um pequeno lampejo de alívio, você percebe não haver comitê de revisão para repreendê-lo. Isso é bom, porque haverá um grande anúncio do Sistema de Reserva Federal amanhã e parece improvável que os preços dos títulos permaneçam iguais. Você não quer gastar 33 minutos preciosos em uma desculpa que provavelmente não precisará.

Sua mente continua vagando pelas explicações que você ouviu na televisão sobre como cada evento se encaixa de forma plausível em sua teoria de mercado. No entanto, fica rapidamente claro que a plausibilidade não pode ajudá-lo aqui — todos os três eventos são plausíveis. A adequação à sua teoria de mercado de animais de estimação não lhe diz como dividir seu tempo. Existe uma lacuna intransponível entre seus 100 minutos, limitados, e sua capacidade ilimitada de explicar como um resultado se encaixa em sua teoria.

E ainda… mesmo em seu estado de espírito incerto, parece que você antecipa os três eventos de maneira diferente, esperando precisar de algumas desculpas mais do que outras. Aqui está a parte fascinante: quando você pensa em algo que torna mais provável que os preços dos títulos subam, você se sente menos propenso a precisar de uma desculpa para os preços dos títulos caírem ou permanecerem os mesmos.

Parece haver uma relação entre o quanto você antecipa cada um dos três resultados e quanto tempo deseja gastar preparando cada desculpa. Claro, essa relação não pode ser quantificada. Você tem 100 minutos para preparar seu discurso, mas não há 100 minutos para dividir entre essas antecipações. Embora você calcule que, se algum resultado específico ocorrer, sua função de utilidade será logarítmica no tempo gasto preparando a desculpa.

Ainda… sua mente continua voltando à ideia de que a antecipação é limitada, ao contrário da desculpabilidade, mas como o tempo para preparar desculpas. Talvez a antecipação deva ser tratada como um recurso conservado, como o dinheiro. Seu primeiro impulso é tentar obter mais antecipação, mas logo você percebe que, mesmo que obtenha mais antecipação, não terá mais tempo para preparar suas desculpas. Seu único curso é alocar seu suprimento limitado de antecipação da melhor maneira possível.

Você tem certeza de que não aprendeu nada parecido em seus cursos de estatística. Eles não lhe disseram o que fazer quando você se sentiu terrivelmente incerto. Eles não lhe disseram o que fazer quando não havia números pequenos para trabalhar. Mesmo que você tente usar números, pode acabar usando qualquer tipo de número — não há nenhuma pista sobre que tipo de matemática usar ou se deve usá-la. Talvez você acabe usando pares de números, números à direita e à esquerda, que você chamaria de DS para Dexter-Sinister. Ou quem sabe o que mais? (Embora você tenha apenas 100 minutos para gastar preparando desculpas.)

Se ao menos houvesse uma arte de focar sua incerteza — de espremer o máximo de antecipação possível em qualquer resultado que realmente aconteça!

Mas como poderíamos chamar uma arte assim? E quais seriam as regras?

21 — O que é evidência?

"A frase — a neve é branca — é verdadeira se e somente se a neve for branca[7]."

— Alfred Tarski

"Dizer do que é, que é, ou do que não é, que não é, é verdadeiro.[8]"

— Aristóteles, Metafísica IV

Se as duas citações anteriores não parecem uma definição suficiente de "verdade", vá para Uma verdade simples. Aqui, falaremos sobre "evidências". (Também planejo discutir crenças factuais, não emoções ou moralidade, conforme distinguimos em Sentimento racional.)

Caminhando pela rua, seus cadarços se desamarram. Pouco tempo depois, por algum motivo estranho, você começa a acreditar que seus cadarços estão desamarrados. A luz deixa o Sol e atinge seus cadarços e ricocheteia; alguns fótons entram nas pupilas de seus olhos e atingem sua retina; a energia dos fótons desencadeia impulsos neurais; os impulsos neurais são transmitidos às áreas de processamento visual do cérebro; e ali a informação ótica é processada e reconstruída em um modelo 3D reconhecido como um cadarço desamarrado. Existe uma sequência de eventos, uma cadeia de causa e efeito, que acontece no mundo e no seu cérebro, que leva você a acreditar no que acredita. O resultado desse processo é um estado mental que reflete o estado real dos seus cadarços reais.

O que são evidências? São eventos interligados, conectados por elos de causa e efeito, relacionados a qualquer coisa que você queira investigar. Por exemplo, se você estiver investigando seus cadarços, a luz que entra em suas pupilas é uma evidência interligada aos seus cadarços. É importante não confundir o termo "emaranhamento" usado na física com o sentido técnico de evidências — aqui, estou falando apenas da interconexão entre duas coisas resultantes em estados correlacionados devido às suas ligações de causa e efeito.

Não é todo tipo de influência que cria o tipo de "emaranhamento" necessário para constituir uma prova. Por exemplo, ter uma máquina que emite um bipe quando se insere números premiados na loteria não é suficiente se essa mesma máquina emite um bipe também quando se insere números perdedores. Da mesma forma, a luz refletida pelos seus sapatos não seria uma evidência útil sobre os seus cadarços se os fótons refletidos terminassem em um mesmo estado físico, independentemente de se os cadarços estivessem amarrados ou desamarrados.

De uma forma abstrata, para um evento ser considerado evidência em uma investigação, ele precisa ocorrer distintamente de modo que esteja interligado aos diferentes possíveis estados do alvo em questão. Tecnicamente falando, isso significa que deve existir uma informação mútua de Shannon entre o evento probatório e o alvo da investigação, em relação ao estado atual de incerteza sobre ambos.

---

7 NT: Tradução livre do texto original em inglês. *The sentence "snow is white" is true if and only if snow is white.*

8 NT: Tradução livre do texto original em inglês. *To say of what is, that it is, or of what is not, that it is not, is true.*

O emaranhamento pode se propagar quando processado corretamente, e é por isso que você precisa dos seus olhos e do seu cérebro. Se os fótons refletidos pelos seus cadarços atingirem uma pedra, a pedra não será alterada significativamente e não refletirá os cadarços de maneira útil. Portanto, não será possível detectar uma diferença entre se os cadarços estavam amarrados ou desamarrados. Isso explica por que as pedras não são consideradas testemunhas úteis no tribunal. Entretanto, um filme fotográfico pode registrar o emaranhamento dos fótons refletidos pelos cadarços, e assim, servir como evidência. Se os seus olhos e cérebro estiverem funcionando adequadamente, você ficará interligado com seus próprios cadarços.

Por isso, os racionalistas valorizam tanto a alegação aparentemente paradoxal de que uma crença só realmente vale a pena se houver a possibilidade de você ser persuadido a acreditar no oposto. Se a sua retina permanecesse no mesmo estado, independentemente da luz que entrasse, você ficaria cego. Alguns sistemas de crenças, em uma tentativa bastante óbvia de se reforçar, afirmam que certas crenças só valem a pena se você acreditar nelas incondicionalmente — não importando o que você veja ou pense. O seu cérebro deve permanecer no mesmo estado independentemente. É daí que surge a expressão "fé cega". Se aquilo em que você acredita não depende do que você vê, você ficou tão cego quanto se tivesse espetado seus próprios olhos.

Se seus olhos e cérebro estiverem funcionando corretamente, suas crenças acabarão se alinhando com os fatos. O pensamento racional gera crenças que são, por si só, evidências. Se você fala a verdade, suas crenças racionais, as quais são evidências por si mesmas, podem servir como evidências para outras pessoas. O emaranhamento pode ser transmitido por meio de cadeias de causa e efeito, e se você fala e outra pessoa ouve, isso também é uma relação causal. Quando você diz "Meus cadarços estão desamarrados" pelo celular, está compartilhando seu emaranhamento dos seus cadarços com um amigo. Portanto, crenças racionais são contagiosas, entre pessoas honestas que acreditam que as outras são honestas. E é por isso que uma alegação de que suas crenças não são contagiosas — que você acredita por razões particulares que não são transmissíveis — é tão suspeita. Se suas crenças estão emaranhadas com a realidade, elas devem ser contagiosas entre as pessoas honestas.

Se você acredita que suas crenças não podem contagiar os outros, isso sugere que elas não são evidências em si e, portanto, não estão emaranhadas com a realidade. É importante que você faça uma reflexão e pare de acreditar nisso.

Na verdade, se você sentir, instintivamente, o que tudo isso significa, você deixará automaticamente de acreditar. Isso porque "minha crença não está emaranhada com a realidade" significa "minha crença não é precisa". Assim que você deixar de acreditar que "'a neve é branca' é verdade", você deve (automaticamente!) deixar de acreditar que "a neve é branca" ou algo está muito errado.

Portanto, explique por que o tipo de processo de pensamento que você utiliza produz sistematicamente crenças que correspondem à realidade. Explique por que você acredita que é racional. Por que você acha que, ao usar processos de pensamento como os que utiliza, as mentes acabam acreditando que "a neve é branca" se e somente se a neve for branca? Se você não acredita que as saídas dos seus processos de pensamento estão interligadas com a realidade, por que acreditar nas saídas dos seus processos de pensamento? É o mesmo ou, pelo menos, deveria ser.

22 — Evidência Científica, Evidência Legal, Evidência Racional

Imagine que seu amigo, o comissário de polícia, confidencialmente, diz que o chefão do crime em sua cidade é Wulky Wilkinsen. Como um racionalista, você pode acreditar nessa afirmação? Bem, se você seguir em frente e insultar Wulky, seria imprudente. Agir como se Wulky fosse um chefão do crime, com uma probabilidade substancialmente maior do que o padrão, é prudente. Portanto, a declaração do comissário de polícia é uma forte evidência bayesiana.

O nosso sistema jurídico não prenderá Wulky com base na declaração do comissário de polícia. Não é admissível como prova legal. Talvez se todas as pessoas acusadas de serem chefes do crime por um comissário de polícia fossem presas, inicialmente muitos chefes do crime seriam pegos, além de algumas poucas pessoas que o comissário de polícia não gostasse. Mas o poder desenfreado atrai a corrupção como o mel atrai as abelhas: com o tempo, você pegaria cada vez menos chefes do crime de verdade (que fariam de tudo para garantir o anonimato) e mais e mais vítimas inocentes.

Isso não quer dizer que a declaração do comissário de polícia não seja uma evidência razoável. Ela ainda tem uma probabilidade desequilibrada e seria imprudente prender Wulky baseado somente nela. Entretanto, em um nível social e visando um objetivo social, definimos deliberadamente a "evidência legal" para incluir somente tipos específicos de evidências, como as próprias observações do comissário de polícia na noite de 4 de abril. Todas as evidências legais deveriam idealmente ser evidências razoáveis, mas o inverso é falso. Exigimos padrões especiais, fortes e adicionais antes de considerarmos as evidências razoáveis como "evidências legais".

Enquanto escrevo esta frase às 20h33, horário do Pacífico, em 18 de agosto de 2007, estou usando meias brancas. Como racionalista, você está autorizado a acreditar na afirmação anterior? Sim. Eu poderia testemunhar isso no tribunal? Sim. É uma afirmação científica? Não, porque não há nenhum experimento que você mesmo possa realizar para verificá-lo. A ciência é composta de generalizações que se aplicam a muitas instâncias particulares, de modo que você pode realizar novos experimentos reais que testam a generalização e, assim, verificar por si mesmo se a generalização é verdadeira, sem ter que confiar na autoridade de ninguém. A ciência é o conhecimento publicamente reproduzível da humanidade.

Assim como o sistema judicial, a ciência também é um processo social que envolve humanos passíveis de erros. Para termos um conjunto de crenças confiáveis, precisamos de regras sociais que incentivem a geração desse conhecimento. Por isso, antes de considerar o conhecimento racional como "científico" e incluí-lo no conjunto de crenças protegidas, impomos padrões rigorosos e adicionais. Como racionalista, é permitido acreditar na existência histórica de Alexandre, o Grande. Temos uma ideia aproximada da Grécia antiga, mas é uma imagem pouco confiável. Embora não possamos verificar tudo por nós mesmos, dependemos de autoridades como Plutarco para obter informações históricas. No entanto, é importante ressaltar que o conhecimento histórico não é o mesmo que o conhecimento científico.

Um racionalista pode acreditar que o Sol nascerá no dia 18 de setembro de 2007? Sim, embora sem certeza, mas essa é a aposta mais segura. (Pedantes: interpretem isso como a rotação e a órbita da Terra, permanecendo aproximadamente constantes em relação ao Sol.) Esta afirmação, enquanto escrevo este ensaio em 18 de agosto de 2007, é uma crença científica?

Negar o adjetivo “científico” a afirmações como “O Sol nascerá em 18 de setembro de 2007” pode parecer perverso. Afinal, se a ciência não pudesse fazer previsões sobre eventos futuros — eventos que ainda não aconteceram — seria inútil e não poderia fazer nenhuma previsão antes do experimento. A previsão de que o Sol nascerá é, definitivamente, uma extrapolação das generalizações científicas. Ela é baseada em modelos do Sistema Solar que podem ser testados por meio de experimentos.

Mas imagine que você está planejando um experimento para verificar a previsão n.º 27 de uma teoria estabelecida Q, em um novo contexto. Você pode não ter razões concretas para suspeitar que a previsão esteja errada; você simplesmente deseja testá-la em um novo ambiente. Parece arriscado afirmar, antes mesmo de realizar o experimento, que existe uma “crença científica” sobre o resultado. Neste caso, podemos falar em uma “previsão convencional” ou “previsão da teoria Q”. Mas, se você já sabe qual é a “crença científica” sobre o resultado, por que se dar ao trabalho de realizar o experimento?

Espero que você esteja começando a entender por que identifico a Ciência com generalizações, em vez da história de qualquer experimento. Um evento histórico acontece uma vez; generalizações se aplicam a muitos eventos. A história não é reproduzível, enquanto as generalizações científicas o são.

A minha definição de “conhecimento científico” é verdadeira? Essa pergunta não está bem formulada. Os padrões específicos que estabelecemos para a ciência são escolhas pragmáticas. Em nenhum lugar no universo está escrito que $p < 0,05$ deve ser o padrão para a publicação científica. Muitos argumentam atualmente que 0,05 é muito fraco e que seria útil reduzi-lo para 0,01 ou 0,001.

Talvez as gerações futuras, agindo com base na teoria de que a ciência é o conhecimento público e repetível da humanidade, rotulem apenas como “científicos” os artigos publicados em um periódico de acesso aberto. Se você cobra pelo acesso ao conhecimento, isso faz parte do conhecimento da humanidade? Podemos confiar em um resultado se as pessoas pagam para criticá-lo? É realmente ciência?

A pergunta “É realmente ciência?” está mal formulada. Um periódico de acesso restrito, que custa US$ 20.000 por ano, é realmente uma evidência bayesiana? Assim como a garantia pessoal do comissário de polícia de que Wulky é o chefão, acredito que devemos responder “Sim”. Contudo, o periódico de acesso restrito deve ser considerado “ciência”? Devemos permitir que ele faça parte do conjunto de crenças especiais e protegidas? Para mim, penso que a ciência seria mais bem servida pelo princípio de que apenas o conhecimento aberto conta como o conjunto de conhecimentos públicos e reproduzíveis da humanidade.

23 — De quanta evidência você precisa?

Anteriormente, defini evidência como “um evento emaranhado, por elos de causa e efeito, com tudo o que você deseja saber” e emaranhado como “ocorrendo de forma diferente para diferentes estados possíveis do alvo”. Então, quanto emaranhamento — quanta evidência — é necessário para sustentar uma crença?

Vamos começar com uma pergunta simples o suficiente para ser respondida pela matemática: Qual é a dificuldade de se ganhar na loteria? Suponha que existam setenta bolas, sorteadas sem reposição, e que são necessários seis números para a vitória. Portanto, existem 131.115.985 combinações vencedoras possíveis, ou seja, um bilhete selecionado aleatoriamente teria uma probabilidade de 1/131.115.985 de ganhar (0,0000007%). Para ganhar na loteria, você precisaria de evidências seletivas suficientes para favorecer uma combinação entre 131.115.984 alternativas.

Imagine que existam testes capazes de discriminar probabilisticamente entre os números vencedores e perdedores da loteria. Por exemplo, você pode colocar uma combinação em uma caixa preta que sempre emite um sinal sonoro se a combinação for vencedora, e tem apenas 1/4 (25%) de chance de apitar se a combinação estiver errada. Usando a abordagem bayesiana, podemos dizer que a razão de verossimilhança é de 4 para 1. Isso significa que a caixa tem quatro vezes mais chances de apitar se a combinação for vencedora, em comparação com a probabilidade de apitar para uma combinação perdedora.

Ainda assim, há muitas possibilidades de combinações. Se você inserir 20 combinações incorretas, a caixa apitará para 5 delas por coincidência (em média). Se você tentar todas as 131.115.985 combinações possíveis, a caixa certamente apitará para a combinação vencedora, mas também para outras 32.778.996 combinações perdedoras (em média).

Assim sendo, essa caixa não fará você ganhar na loteria, mas já é alguma coisa. Se você a utilizar, suas chances de ganhar aumentariam de 1 em 131.115.985 para 1 em 32.778.997. Dessa forma, você terá avançado na busca pelo seu objetivo, a verdade, em meio ao vasto espaço de possibilidades.

Digamos que você possa usar outra caixa preta para testar as combinações duas vezes, independentemente. Ambas as caixas irão com certeza apitar para o bilhete premiado. No entanto, a chance de uma caixa apitar para uma combinação perdedora é de 1/4, independente de cada caixa. Portanto, a chance de ambas as caixas apitarem para uma combinação perdedora é de 1/16. Podemos afirmar que a evidência cumulativa de dois testes independentes tem uma razão de verossimilhança de 16:1. O número de bilhetes de loteria perdedores que passarão em ambos os testes será, em média, de 8.194.749.

Considerando que existem 131.115.985 bilhetes possíveis na loteria, podemos supor que precisemos de evidências tão fortes quanto 131.115.985 para 1 - um evento ou série de eventos com 131.115.985 vezes mais probabilidade de ocorrer para uma combinação vencedora do que para uma combinação perdedora. No entanto, essa quantidade de evidências seria suficiente apenas para lhe dar uma chance uniforme de ganhar na loteria. Por quê? Porque se você aplicar um filtro dessa potência elevada a 131 milhões de bilhetes perdedores, haverá, em média, um bilhete perdedor que passará pelo filtro. O bilhete premiado também passará pelo filtro. Portanto, você terá dois bilhetes que passaram pelo filtro, mas apenas um deles será o vencedor. Se você puder comprar apenas um bilhete, suas chances de ganhar serão de cinquenta por cento.

Uma maneira mais clara de entender o problema é a seguinte: inicialmente, há um bilhete vencedor e 131.115.984 bilhetes perdedores, o que significa que suas chances de ganhar são de 1 em 131.115.984. Ao usar uma única caixa preta, a chance da caixa apitar é de 1 para o bilhete premiado e de 0,25 para um bilhete perdedor. Portanto, multiplicamos 1 em 131.115.984 por 1 em 0,25 e obtemos 1 em 32.778.996. Se adicionarmos outra caixa de evidência, as chances serão multiplicadas novamente por 1 em 0,25, resultando em 1 bilhete vencedor para cada 8.194.749 bilhetes perdedores.

É conveniente medir as evidências em bits — não como os bits de um disco rígido, mas os bits dos matemáticos, os quais são conceitualmente diferentes. Os bits do matemático são os logaritmos, base 1/2, das probabilidades. Por exemplo, se houver quatro resultados possíveis A, B, C e D, cujas probabilidades são 50%, 25%, 12,5% e 12,5%, e eu disser que o resultado foi "D", então transmiti três bits de informação para você, porque eu o informei sobre um resultado cuja probabilidade era 1/8.

Com 131.115.984 possibilidades de bilhetes, essa quantidade é um pouco menor que 2 elevados à 27ª potência. Assim, se 14 caixas de evidência ou 28 bits de informação fossem adicionados — um evento seria 268.435.456:1 mais provável de ocorrer se a hipótese de ter o bilhete vencedor for verdadeira do que se for falsa — as chances de ganhar mudariam de 1:131.115.984 para 268.435.456:131.115.984, o que pode ser reduzido para 2:1. Ter probabilidades de 2 para 1 significa que há duas chances de ganhar para cada chance de perder, resultando em uma probabilidade de ganhar de 2/3 com 28 bits de evidência. Acrescentar uma caixa, adicionando 2 bits de evidência, aumentaria as chances para 8:1, e mais duas caixas aumentariam as chances para 128:1.

Dessa forma, se o seu objetivo é ter uma forte convicção de que irá ganhar na loteria —definida arbitrariamente como ter menos de 1% de chance de estar errado, então você precisará de 34 bits de evidência sobre a combinação vencedora para alcançar esse objetivo.

Em geral, as regras para avaliar quantas evidências são necessárias seguem um padrão semelhante: quanto maior o espaço de possibilidades em que se encontra a hipótese, ou quanto mais improvável a hipótese parecer a priori em comparação com outras hipóteses, ou ainda, quanto mais confiante se deseja estar, maior será a quantidade de evidências necessárias.

As regras não podem ser desafiadas; crenças precisas não podem ser formadas com evidências inadequadas. Imagine que você tenha 10 caixas alinhadas e comece a testar combinações nelas. Não é aceitável parar na primeira combinação que faz as 10 caixas apitarem e dizer: "Mas a chance de isso acontecer com uma combinação perdedora é de um milhão para um! Ignorarei as regras Bayesianas e parar aqui". Considerando o espaço de possibilidades e a improbabilidade inicial, você tirou uma conclusão muito forte com base em evidências insuficientes. Isso não é uma burocracia sem sentido; é matemática.

Claro, você ainda pode acreditar com base em evidências insuficientes, se quiser; mas não poderá acreditar com precisão. É como tentar dirigir um carro sem combustível, porque você não acredita na ideia tola de que precisa de combustível para se locomover. Seria muito mais divertido e muito mais barato se simplesmente revogássemos a lei que diz que carros precisam de combustível. Não seria ótimo para todos? Bem, você pode tentar, se quiser. Você pode até fechar os olhos e fingir que o carro está se movendo. Mas, para acreditar com precisão, é necessário o combustível das evidências e, quanto mais longe você quiser ir, mais combustível precisará.

24 — Arrogância de Einstein

Em 1919, Sir Arthur Eddington liderou expedições ao Brasil e à Ilha do Príncipe para observar eclipses solares e testar uma previsão experimental da nova teoria da Relatividade Geral de Einstein. Quando um jornalista perguntou a Einstein o que ele faria se as observações de Eddington não correspondessem à sua teoria, Einstein respondeu: "Então, eu sentiria pena do bom Deus. A teoria está correta."

Essa afirmação parece bastante imprudente e desafia o tropo da Racionalidade Tradicional de que o experimento acima de tudo é soberano. Parece que Einstein está possuído por uma arrogância tão grande que se recusaria a se submeter à resposta da natureza, como os cientistas devem fazer. Mas quem pode saber se a teoria está correta antes do teste experimental?

No entanto, Einstein acabou por estar certo. Tento evitar criticar as pessoas quando elas estão certas. Se elas genuinamente merecem críticas, não precisarei esperar muito por uma ocasião em que elas estejam erradas.

Einstein pode não ter sido tão temerário quanto parecia…

Para atribuir mais de 50% de probabilidade ao candidato correto de um conjunto de 100.000.000 de hipóteses possíveis, você precisa de pelo menos 27 bits de evidência (ou por aí). Você não pode esperar encontrar o candidato correto sem testes tão fortes, porque testes menores renderão mais de um candidato que passa em todos os testes. Se você tentar aplicar um teste que tenha apenas uma chance de um milhão para um falso positivo (~ 20 bits), você acabará com cem candidatos. Apenas encontrar a resposta certa, em um grande espaço de possibilidades, requer uma abundância de evidências.

A Racionalidade Tradicional enfatiza a justificação: "Se você quer me convencer de X, você tem que me apresentar Y quantidade de evidências." Eu mesmo costumo usar essa frase sempre que digo algo como: "Para justificar a crença nesta proposição, com mais de 99% de probabilidade, são necessários 34 bits de evidência". Ou, "Para atribuir mais de 50% de probabilidade à sua hipótese, você precisa de 27 bits de evidência". A frase tradicional implica que você começa com um palpite, ou alguma linha particular de raciocínio que o leva a uma hipótese sugerida, e então você tem que reunir "evidências" para confirmá-la —para convencer a comunidade científica ou justificar dizendo que você acredita em seu palpite.

Mas, de uma perspectiva bayesiana, você precisa de uma quantidade de evidências aproximadamente equivalente à complexidade da hipótese apenas para localizar a hipótese no espaço teórico. Não é uma questão de justificar nada a ninguém. Se houver cem milhões de alternativas, você precisa de pelo menos 27 bits de evidência apenas para focar sua atenção exclusivamente na resposta correta.

Isso é verdade, mesmo que você chame sua suposição de "palpite" ou "intuição". Intuições são processos reais em um cérebro real. Se o seu cérebro não tiver pelo menos 10 bits de evidências bayesianas genuinamente conectadas para analisar, ele não poderá destacar uma hipótese correta de 10 bits para sua atenção — consciente, subconsciente, ou qualquer outra. Os processos subconscientes não podem encontrar um alvo em um milhão usando apenas 19 bits de conexão, assim como os processos conscientes. Palpites podem ser misteriosos para os curiosos, mas não podem violar as leis da física.

Você vê para onde isso está indo: no momento em que formulou a hipótese pela primeira vez, a primeira vez que as equações surgiram em sua cabeça, Einstein já deveria ter em sua posse evidências observacionais suficientes para destacar as complexas equações da Relatividade Geral para sua atenção exclusiva. Caso contrário, ele não poderia tê-las acertado.

Agora, qual é a probabilidade de Einstein ter exatamente evidências observacionais suficientes para elevar a Relatividade Geral ao nível de sua atenção, mas apenas justificar atribuir a ela uma probabilidade de 55%? Suponha que a Relatividade Geral seja uma hipótese de 29,3 bits. Qual a probabilidade de Einstein tropeçar em exatamente 29,5 bits de evidência no curso de sua leitura de física?

Não é provável! Se Einstein tinha evidências observacionais suficientes para identificar as equações corretas da Relatividade Geral em primeiro lugar, então ele provavelmente tinha evidências suficientes para ter certeza de que a Relatividade Geral era verdadeira.

Na verdade, uma vez que o cérebro humano não é um processador de informações perfeitamente eficiente, Einstein provavelmente tinha muito mais evidências do que, em princípio, seria necessário para um Bayesiano perfeito atribuir confiança maciça à Relatividade Geral.

“Então, eu sentiria pena do bom Deus; a teoria está correta”. Não soa tão terrível quando você olha dessa perspectiva. E lembre-se de que a Relatividade Geral estava correta, em todo aquele vasto espaço de possibilidades.

25 — Navalha de Ocam

Quanto mais complexa for uma explicação, mais evidências são necessárias para encontrá-la no espaço das crenças. Na Racionalidade Tradicional, isso costuma ser formulado erroneamente como "Quanto mais complexa é uma proposição, mais evidências são necessárias para defendê-la". Como podemos medir a complexidade de uma explicação? Como podemos determinar a quantidade de evidências necessárias?

A Navalha de Ocam é frequentemente definida como "a explicação mais simples que se ajusta aos fatos". Robert Heinlein respondeu que a explicação mais simples é "a senhora que mora no fim da rua é uma bruxa; foi ela que fez isso".

Observa-se que o comprimento de uma frase em inglês não é uma boa maneira de medir a complexidade. E "encaixar" os fatos, deixando meramente de proibi-los, é insuficiente.

Por que exatamente o comprimento de uma frase em inglês é uma medida pobre de complexidade? Porque, quando você fala uma frase em voz alta, está usando rótulos para conceitos que o ouvinte compartilha — o receptor já armazenou a complexidade neles. Suponha que abreviamos toda a frase de Heinlein como "Asqmnfdréubfeqfi!" para que toda a explicação possa ser transmitida em uma palavra; melhor ainda, daremos a ela um rótulo curto e arbitrário como "Fnord!" Isso reduz a complexidade? Não, porque você tem que dizer ao ouvinte com antecedência que "Asqmnfdréubfeqfi!" significa "A senhora que mora no fim da rua é uma bruxa; foi ela que fez isso". Bruxa, em si, é um rótulo para algumas afirmações extraordinárias, só porque todos sabemos o que isso significa, não significa que o conceito seja simples.

Um enorme raio de eletricidade sai do céu e atinge alguma coisa, e o povo da tribo nórdica diz: "Talvez um agente realmente poderoso estivesse com raiva e atirou um raio". O cérebro humano é o artefato mais complexo do universo conhecido. Se a raiva parece simples, é porque não vemos todos os circuitos neurais que implementam a emoção. Imagine tentar explicar por que o Saturday Night Live[9] é engraçado, para uma espécie alienígena sem senso de humor. Mas não se sinta superior; você mesmo não tem senso de fnord.) A complexidade da raiva e, na verdade, a complexidade da inteligência, foi encoberta pelos humanos que levantaram a hipótese de Thor, o agente do trovão.

Para um ser humano, explicar as equações de Maxwell leva muito mais tempo do que explicar as de Thor. Os humanos não possuem um vocabulário inato para cálculo, da mesma forma que possuímos um vocabulário inato para expressar a emoção da raiva. Antes de começar a falar sobre eletricidade, é preciso primeiro explicar a linguagem, a metalinguagem e os próprios conceitos matemáticos.

No entanto, há uma maneira na qual as equações de Maxwell podem ser mais simples do que o cérebro humano ou Thor, o agente do trovão: a programação de software.

Escrever um software que simule as equações de Maxwell é muito mais fácil do que desenvolver um software que emule uma mente emocional inteligente como a de Thor.

9 NT: Programa de comédia de televisão norte-americano exibido aos sábados à noite.

O formalismo da indução de Solomonoff mede a complexidade da descrição pelo comprimento do software mais curto que produz essa descrição como saída. Para falar sobre o menor software que realiza uma determinada tarefa, é necessário especificar um espaço de softwares, requerendo uma linguagem de programação e um intérprete. A indução Solomonoff usa máquinas de Turing, ou mais especificamente, cadeias de bits que especificam máquinas de Turing. Mas, e se você não quiser usar máquinas de Turing? Não se preocupe, há apenas uma penalidade de complexidade constante para projetar sua própria máquina de Turing universal, que pode interpretar qualquer código que você fornecer a ela, não importa qual linguagem de programação você escolha. Diferentes formalismos indutivos têm um fator constante de pior caso relativo entre si, correspondente ao tamanho de um intérprete universal para aquele formalismo.

Nas melhores versões (na minha humilde opinião) da indução de Solomonoff, o software não produz uma previsão determinística, mas atribui probabilidades a strings. Por exemplo, podemos escrever um programa para explicar uma moeda honesta escrevendo um programa que atribui probabilidades iguais a todas as duas elevadas à potência N**,** cadeias de comprimento N. Essa é a abordagem da indução de Solomonoff para ajustar os dados observados. Quanto maior a probabilidade que um programa atribui aos dados observados, melhor esse programa ajusta os dados. E as probabilidades devem somar 1, portanto, para um programa ajustar melhor uma possibilidade, ele deve roubar a massa de probabilidade de alguma outra possibilidade que então se ajustará muito pior. Não existe moeda completamente justa que atribua 100% de probabilidade para cara e 100% de probabilidade para coroa.

Como podemos compensar o ajuste dos dados em relação à complexidade do programa? Se você ignorar as penalidades de complexidade e pensar apenas no ajuste, sempre escolherá programas que afirmam prever os dados de forma determinística, atribuindo-lhes 100% de probabilidade. Se a moeda mostrar CAcocoCACAco, então o programa que afirma que a moeda foi corrigida para mostrar CAcocoCACAco ajusta os dados observados 64 vezes melhor do que o programa que afirma que a moeda é honesta. Por outro lado, se você ignorar o ajuste e considerar apenas a complexidade, a hipótese de moeda honesta sempre parecerá mais simples do que qualquer outra hipótese, mesmo que a moeda mostre sequências como CAcoCACAco-CA-CACAcoCACACACAcoCACACACACAco. Na verdade, a moeda honesta é mais simples e se ajusta a esses dados exatamente tão bem quanto a qualquer outra sequência de 20 caras ou coroas — nem mais, nem menos —mas há outra hipótese que não parece muito complicada e que se ajusta muito melhor aos dados.

Se permitirmos que um programa armazene mais um bit binário de informação, ele conseguirá reduzir pela metade o espaço de possibilidades e, portanto, atribuir o dobro de probabilidade a todos os pontos no espaço restante. Isso sugere que um bit de complexidade do programa deve custar pelo menos um “fator de ganho de dois” no ajuste. Se tentarmos projetar um software que armazene explicitamente um resultado como CAcocoCACAco, os seis bits que perdemos em complexidade devem destruir toda a plausibilidade obtida por uma melhoria de ajuste de 64 vezes. Caso contrário, mais cedo ou mais tarde, decidiremos que todas as moedas justas são fixas.

A menos que o programa seja inteligente e comprima os dados, não adianta apenas mover um bit dos dados para a descrição do programa.

A forma como a indução de Solomonoff funciona para prever sequências é a seguinte: você soma todos os softwares permitidos (se algum programa for permitido, a indução de Solomonoff se torna incomputável), e cada programa tem uma probabilidade anterior de (1/2) elevado à potência do seu comprimento em bits. Cada programa é ainda mais ponderado pelo ajuste a todos os dados observados até agora, fornecendo assim uma mistura ponderada de especialistas que podem prever bits futuros.

O formalismo do Comprimento Mínimo da Mensagem é quase equivalente à indução de Solomonoff. Você envia uma string descrevendo um código e, em seguida, envia uma string descrevendo os dados desse código. Qualquer explicação que leve à mensagem total mais curta é a melhor. Se você pensar no conjunto de códigos permitidos como um espaço de softwares e na linguagem de descrição de código como uma máquina universal, o Comprimento Mínimo da Mensagem é quase equivalente à indução de Solomonoff (quase porque escolhe o programa mais curto, em vez de resumir todos os programas).

Isso nos permite ver claramente o problema de usar a frase “A senhora da rua é uma bruxa; ela é culpada” para explicar o padrão na sequência 0101010101. Se você estiver enviando uma mensagem para um amigo, tentando descrever a sequência que observou, deverá dizer: “A senhora da rua é uma bruxa; ela fez a sequência sair 0101010101.” A acusação de bruxaria não permitiu que você encurtasse o resto da mensagem. Você ainda teria que descrever, em todos os detalhes, os dados que a bruxaria causou.

A feitiçaria pode se encaixar em nossas observações no sentido de permiti-las qualitativamente, mas é porque a bruxaria permite tudo, assim como dizer “flogisto!” Então, mesmo após dizer “bruxa”, você ainda precisa descrever todos os dados observados em detalhes. Você não comprimiu o comprimento total da mensagem descrevendo suas observações ao transmitir a mensagem sobre bruxaria; você simplesmente adicionou um prólogo inútil, aumentando o comprimento total.

A verdadeira dissimulação estava escondida na palavra “isso” de “Uma bruxa fez isso”. Uma bruxa fez o quê?

Claro, graças ao viés retrospectivo, ancoragem, explicações falsas, causalidade falsa, viés positivo e cognição motivada, pode parecer muito óbvio que, se uma mulher é uma bruxa, é claro que ela faria a moeda 0101010101 aparecer. Mas chegaremos a isso em breve.

26 — Sua força como um racionalista

O seguinte aconteceu comigo em uma sala de bate-papo do IRC, já há bastante tempo, visto que eu ainda estava entrando em salas de bate-papo do IRC nessa época. O tempo embaçou um pouco a minha memória e meu relato pode ser impreciso.

Então, lá estava eu, em uma sala de bate-papo IRC, quando alguém relata que um amigo dele precisa de conselhos médicos. O amigo dele diz que tem sentido dores repentinas no peito, então ele chamou uma ambulância, e a ambulância chegou, mas os paramédicos disseram que não era nada e foram embora, e agora as dores no peito estão piorando. O que o amigo dele deveria fazer?

Essa história me deixou confuso. Lembro-me de ter lido sobre pessoas em situação de rua em Nova York que chamavam ambulâncias apenas para serem levadas a um lugar quente, e como os paramédicos sempre tinham que levá-los à sala de emergência, mesmo pela 27ª vez. Porque, se eles não o fizessem, a empresa de ambulâncias poderia ser processada por muito dinheiro. Da mesma forma, as salas de emergência têm a obrigação legal de tratar qualquer pessoa, independentemente da capacidade de pagamento. (E o hospital absorve os custos, que são enormes, então os hospitais estão fechando suas emergências… Isso faz a gente se perguntar qual é o sentido de ter economistas se nós simplesmente vamos ignorá-los. Então, eu não consegui entender muito bem como os eventos descritos puderam ter acontecido. Qualquer pessoa relatando dores repentinas no peito deveria ter sido levada imediatamente por uma ambulância.

E foi aqui que fracassei como racionalista. Lembrei-me de várias ocasiões em que meu médico não demonstrou nenhum pânico ao ouvir relatos de sintomas que pareciam muito alarmantes para mim. E a instituição médica estava sempre certa. Sempre. Eu mesmo tive dores no peito em um momento, e o médico pacientemente me explicou que eu estava descrevendo dor muscular no peito, não um ataque cardíaco. Então, eu disse no canal IRC: “Bem, se os paramédicos disseram ao seu amigo que não era nada, realmente não deve ser nada — eles o teriam arrastado se houvesse a menor chance de problemas sérios”.

Assim, consegui explicar a história dentro do meu modelo existente, embora a adaptação ainda parecesse um pouco forçada…

Mais tarde, o sujeito volta para a sala de chat do IRC e diz que seu amigo inventou toda a história. Evidentemente, este não era um de seus amigos mais confiáveis.

Eu deveria ter percebido, talvez, que um amigo desconhecido de um conhecido em um canal IRC poderia ser menos confiável do que um artigo publicado em revista. Infelizmente, a crença é mais fácil do que a descrença; acreditamos instintivamente, mas a descrença exige um esforço consciente. [1]

Assim, em vez disso, por meio de grande esforço, forcei meu modelo de realidade a explicar uma anomalia que nunca aconteceu. E eu sabia o quão constrangedor isso era. Eu sabia que a utilidade de um modelo não é o que ele pode explicar, mas o que ele não pode. Uma hipótese que não proíbe nada, permite tudo e, portanto, falha em limitar a antecipação.

Sua força como racionalista é sua capacidade de ficar mais confuso com a ficção do que com a realidade. Se você é igualmente bom em explicar qualquer resultado, você não tem conhecimento algum.

Todos somos fracos, ocasionalmente; a parte triste é que eu poderia ter sido mais forte. Eu tinha todas as informações necessárias para chegar à resposta correta, até percebi o problema e, em seguida, o ignorei. Minha sensação de confusão foi uma pista, e eu joguei minha pista fora.

Eu deveria ter prestado mais atenção a essa sensação de "ainda parece um pouco forçado". É um dos sentimentos mais importantes que um buscador da verdade pode ter, uma parte de sua força como racionalista. É uma falha de projeto na cognição humana que essa sensação se manifeste como uma tensão silenciosa no fundo da mente, em vez de uma sirene de alarme gritante e um letreiro de néon brilhante que diz:

OU O SEU MODELO ESTÁ ERRADO, OU ESSA HISTÓRIA ESTÁ ERRADA.

Referências

1. *Daniel T. Gilbert, Romin W. Tafarodi, and Patrick S. Malone, "You Can't Not Believe Everything You Read," Journal of Personality and Social Psychology 65 (2 1993): 221–233, doi:10.1037/0022-3514.65.2.221.*

27 — Ausência de evidência é evidência de ausência

De *Rational Choice in an Uncertain World* (Uma escolha racional em um mundo incerto), de Robyn Dawes: [1]

Esse ajuste post-hoc de evidências à hipótese, na verdade, está relacionado a um capítulo doloroso da história dos Estados Unidos: o internamento de nipo-americanos no início da Segunda Guerra Mundial. Quando o governador da Califórnia, Earl Warren, depôs perante uma audiência do Congresso em San Francisco, em 21 de fevereiro de 1942, um dos questionadores apontou que, até aquele momento, não havia ocorrido sabotagem ou qualquer outro tipo de espionagem por parte dos nipo-americanos. Warren respondeu: "Acredito que essa ausência [de atividades subversivas] é o sinal mais ameaçador em toda a nossa situação. Isso me convence mais do que qualquer outro fator que a sabotagem que devemos enfrentar, as atividades da Quinta Coluna devem estar cronometradas exatamente como o ataque a Pearl Harbor. … Acredito que estamos sendo iludidos por uma falsa sensação de segurança."[10]

Considerando o argumento de Warren sob a perspectiva bayesiana, podemos afirmar que, ao observarmos evidências, as hipóteses que atribuem uma maior probabilidade a essa evidência ganham força em relação às hipóteses que atribuem uma probabilidade menor a ela. Esse fenômeno é conhecido como probabilidade relativa e pode levar a uma perda de probabilidade em relação a outras hipóteses, se estas atribuírem uma probabilidade ainda maior à evidência. Ou seja, mesmo que uma alta probabilidade seja atribuída à evidência, isso não garante que essa hipótese seja a mais provável, caso haja outra hipótese que atribua uma probabilidade ainda maior.

Warren parece estar defendendo que, como não há sinais de sabotagem, isso confirma a presença de uma Quinta Coluna. No entanto, é possível argumentar que uma Quinta Coluna poderia estar atrasando sua ação de sabotagem. Contudo, é mais provável que a ausência de evidências da presença de uma Quinta Coluna signifique a ausência de sabotagem.

Consideremos E como a observação de sabotagem e ¬E como a observação de não sabotagem. H1 representa a hipótese de uma Quinta Coluna nipo-americana e H_2 a hipótese de que não existe uma Quinta Coluna. A probabilidade condicional P(E|H), ou "E dado H", representa a confiança que teríamos em ver a evidência E se assumíssemos que a hipótese H fosse verdadeira.

Independentemente da probabilidade de que uma Quinta Coluna não faça sabotagem, a probabilidade P ($\neg E|H_1$), não será tão grande quanto a probabilidade de que não haja sabotagem, dado que não há Quinta Coluna, a probabilidade P ($\neg E|H_2$). Portanto, observar a falta de sabotagem aumenta a probabilidade de que não exista uma Quinta Coluna.

A falta de sabotagem não prova que não exista uma Quinta Coluna. A ausência de prova não é prova

10 NT: Tradução livre do texto original em inglês. *In fact, this post-hoc fitting of evidence to hypothesis was involved in a most grievous chapter in United States history: the internment of Japanese-Americans at the beginning of the Second World War. When California governor Earl Warren testified before a congressional hearing in San Francisco on February 21, 1942, a questioner pointed out that there had been no sabotage or any other type of espionage by the Japanese-Americans up to that time. Warren responded, "I take the view that this lack [of subversive activity] is the most ominous sign in our whole situation. It convinces me more than perhaps any other factor that the sabotage we are to get, the Fifth Column activities are to get, are timed just like Pearl Harbor was timed… I believe we are just being lulled into a false sense of security.*

de ausência. Na lógica, (A ⇒ B), leia-se “A implica B,” não é equivalente a (¬A ⇒ ¬B), leia-se “não-A implica não-B.”

Na teoria da probabilidade, a ausência de evidência é sempre considerada evidência de ausência. Considere um evento binário E, e uma hipótese H. Se a ocorrência de E aumenta a probabilidade de H, ou seja, se P(H|E) > P(H), então a não observação de E diminuirá a probabilidade de H, o que é representado por P(H | ¬ E) < P(H). A probabilidade P(H) é uma média ponderada de P(H | E) e P(H | ¬ E) e, portanto, sempre se encontrará entre esses dois valores. Se algum desses conceitos parecer confuso, recomendo consultar o texto “Uma Explicação Intuitiva do Teorema de Bayes” para obter mais esclarecimentos.

Na maioria das situações da realidade, uma causa pode não apresentar sinais confiáveis, mas a ausência da causa é ainda menos provável de gerar esses sinais. A falta de uma observação pode ser uma evidência forte ou fraca de ausência, dependendo da probabilidade de a causa gerar a observação. Quando a ausência de uma observação é apenas levemente permitida (mesmo que a hipótese alternativa não o permita), a evidência é muito fraca de ausência (mas ainda é uma evidência). Isso é conhecido como a falácia das “lacunas no registro fóssil” —fósseis raramente se formam, então é inútil se preocupar com a ausência de uma observação fracamente permitida quando muitas observações positivas fortes já foram registradas. No entanto, se não houver observações positivas, é hora de se preocupar, e é aí que surge o Paradoxo de Fermi.

Um racionalista forte é aquele que consegue se confundir mais facilmente com a ficção do que com a realidade. Se você consegue explicar qualquer resultado com igual facilidade, isso significa que não possui conhecimento algum. A força de um modelo não está em sua capacidade de explicação, mas sim em suas limitações, pois apenas as restrições definem o que pode ser antecipado. Se você não perceber quando seu modelo torna a evidência improvável, pode ser que ele seja inválido, ou até mesmo que não haja nenhuma evidência disponível. Nesse caso, não haverá nenhum cérebro ou olho capaz de processar a informação.

Referências

1. *Robyn M. Dawes, Rational Choice in An Uncertain World, 1st ed., ed. Jerome Kagan (San Diego, CA: Harcourt Brace Jovanovich, 1988), 250–251.*

28 — Conservação da evidência esperada

O padre Friedrich Spee von Langenfeld, que ouviu confissões de bruxas condenadas, escreveu em 1631 o *Cautio Criminalis* (prudência em casos criminais), onde satiricamente descreveu a árvore de decisão usada para condenar bruxas acusadas: se a bruxa levava uma vida má e imprópria, ela era culpada; se levava uma vida boa e adequada, isso também era uma prova, pois as bruxas dissimulam e tentam parecer especialmente virtuosas. Depois que a mulher era presa, se estava com medo, isso provava sua culpa; se não estava com medo, isso também provava sua culpa, já que as bruxas normalmente fingem inocência e adotam uma atitude corajosa. Se ouvisse uma denúncia de bruxaria contra ela, poderia fugir ou permanecer; se fugisse, isso provava sua culpa; se ficasse, o diabo a havia impedido de fugir.

Como confessor de várias bruxas condenadas, Spee tinha acesso privilegiado a cada detalhe das acusações, e percebeu que não importava o que a bruxa dissesse ou fizesse, tudo era considerado uma prova contra ela. Cada ramo da árvore de decisão levava à condenação, de modo que, em qualquer caso individual, você só ouviria um lado do dilema. Essa é uma das razões pelas quais os cientistas escrevem suas previsões experimentais com antecedência.

No entanto, você não pode ter as duas coisas, e isso não se trata apenas de justiça, mas de uma questão de teoria da probabilidade. A regra que diz que "a falta de evidência é evidência da falta" é um caso particular de uma lei mais geral que eu chamaria de Conservação da Evidência Esperada. Essa lei afirma que a expectativa da probabilidade posterior, após a observação da evidência, deve ser igual à probabilidade anterior.

$$P(H) = P(H, E) + P(H, \neg E)$$

$$P(H) = P(H|E) \times P(E) + P(H|\neg E) \times P(\neg E)$$

Portanto, para cada expectativa de evidência, há uma expectativa igual e oposta de contraprova.

Se você espera uma forte probabilidade de encontrar evidências fracas em uma direção, isso deve ser equilibrado por uma expectativa fraca de encontrar evidências fortes na outra direção. Se você está muito confiante em sua teoria e, portanto, espera ver um resultado que confirme sua hipótese, isso pode fornecer apenas um pequeno incremento em sua crença (já está próxima de 1); mas a falha inesperada de sua previsão deve causar um grande golpe em sua confiança. Em média, você deve esperar estar exatamente tão confiante quanto antes. Da mesma forma, a mera expectativa de encontrar evidências — antes mesmo de você vê-las —não deve mudar suas crenças anteriores. (Novamente, se isso não for intuitivamente óbvio, consulte Uma Explicação Intuitiva do teorema de Bayes.)

Então, se você afirma que "sem sabotagem" é evidência para a existência de uma Quinta Coluna nipo-americana, você deve, inversamente, sustentar que ver sabotagem seria um argumento contra uma Quinta Coluna. Se você afirma que "uma vida boa e adequada" é evidência de que uma mulher é uma bruxa, então uma vida má e imprópria deve ser evidência de que ela não é uma bruxa. Se você argumenta que Deus, para testar a fé da humanidade, se recusa a revelar Sua existência, então os milagres descritos na Bíblia devem argumentar contra a existência de Deus.

Perceba haver algo que não soa natural. É preciso prestar atenção a essa sensação de que algo soa um pouco forçado, a tensão silenciosa que se instala no fundo da mente. Isso é importante.

Um verdadeiro bayesiano não pode procurar evidências que confirmem uma teoria. Não há nenhum plano, estratégia inteligente ou dispositivo astuto que permita aumentar legitimamente a confiança em uma proposição fixa, em média. A única possibilidade é buscar evidências para testar uma teoria, e não para confirmá-la.

Essa compreensão pode trazer grande alívio para a sua mente. Não há necessidade de se preocupar com a interpretação de todos os possíveis resultados experimentais para confirmar a sua teoria. Não é necessário planejar estratégias para fazer qualquer evidência confirmar a sua teoria, porque você sabe que, para cada expectativa de evidência, há uma expectativa igual e oposta de contra-evidência. Qualquer tentativa de enfraquecer a contra-evidência de uma possível observação "anormal" só poderá ser feita enfraquecendo o suporte de uma observação "normal" em um grau precisamente igual e oposto. É um jogo de soma zero. Não importa como você conspira, argumenta ou cria estratégias, você não pode esperar que o resultado do jogo mude suas crenças (em média) em uma direção específica.

Você pode se permitir sentar e relaxar enquanto aguarda as evidências chegarem.

… A psicologia humana é tão problemática.

29 — A visão do passado desvaloriza a ciência

Este ensaio é baseado em um trecho de Exploring Social Psychology (Explorando a psicologia social), de Meyers [1]; vale a pena ler o trecho completo.

Cullen Murphy, editor do *The Atlantic*, disse que as ciências sociais não apresentam "nenhuma ideia ou conclusão que não possa ser encontrada em [qualquer] enciclopédia de citações… Dia após dia, os cientistas sociais saem para o mundo. Dia após dia, eles descobrem que o comportamento das pessoas é exatamente o que você esperaria."

Claro, a "expectativa" é toda retrospectiva. (Viés retrospectivo: sujeitos que sabem a resposta real a uma pergunta atribuem probabilidades muito mais altas do que "teriam" adivinhado para essa resposta, em comparação com sujeitos que devem adivinhar sem saber a resposta.)

O historiador Arthur Schlesinger Jr. descartou os estudos científicos das experiências dos soldados da Segunda Guerra Mundial como "demonstrações ponderosas" do senso comum. Por exemplo:

1. Soldados com maior escolaridade sofreram mais problemas de adaptação do que soldados com menor escolaridade. (Os intelectuais estavam menos preparados para o estresse da batalha do que os menos instruídos.)
2. Os soldados do sul lidaram melhor com o clima quente da Ilha dos Mares do Sul do que os soldados do norte. (Os sulistas estão mais acostumados com o clima quente.)
3. Soldados brancos estavam mais ansiosos para serem promovidos a suboficiais do que soldados negros. (Anos de opressão prejudicaram a motivação para alcançar uma promoção.)
4. Os negros do sul preferiam os oficiais brancos do sul aos do norte. (Oficiais do sul eram mais experientes e habilidosos em interagir com os negros.)
5. Enquanto a luta continuava, os soldados estavam mais ansiosos para voltar para casa do que após o fim da guerra. (Durante a luta, os soldados sabiam correrem risco de vida.)

Quantas dessas descobertas você acha que poderia ter previsto com antecedência? Três em cinco? Quatro em cinco? Há algum caso em que você teria previsto o oposto — em que seu modelo foi atingido? Reserve um momento para pensar antes de continuar.

Nesta demonstração (de Paul Lazarsfeld por meio de Meyers), todas as descobertas acima são o oposto do que foi encontrado de fato. [2] Quantas vezes você pensou que seu modelo foi atingido? Quantas vezes você admitiu que teria errado? É assim que seu modelo realmente era bom. A medida de sua força como racionalista é sua capacidade de se confundir mais com a ficção do que com a realidade.

Exceto se eu inverta novamente os resultados. O que você acha?

Seus processos de pensamento agora, quando você realmente não sabe a resposta, parecem diferentes dos processos de pensamento que você usou para racionalizar um dos lados da resposta "conhecida"?

Daphna Baratos apresentou aos estudantes universitários pares de supostas descobertas, uma verdadeira ("Em tempos prósperos, as pessoas gastam uma parte maior de sua renda do que durante uma recessão") e outra que é o oposto da verdade. [3] Em ambos os lados do par, os alunos avaliaram a suposta descoberta como o que eles "teriam previsto". Isso é um viés retrospectivo perfeitamente comum.

O que leva as pessoas a pensar que não precisam da ciência, porque "poderiam ter previsto" isso.

(Exatamente como você esperaria, certo?)

A retrospectiva nos levará a subestimar sistematicamente a surpresa das descobertas científicas, especialmente aquelas que entendemos — aquelas que parecem reais para nós, aquelas que podemos adaptar aos nossos modelos do mundo. Se você entende de neurologia ou física e lê notícias sobre esses tópicos, subestima provavelmente a surpresa das descobertas nessas áreas também. Isso desvaloriza injustamente a contribuição dos pesquisadores; e pior, impedirá que você perceba quando estiver vendo evidências que não se encaixam no que você realmente esperava.

Precisamos fazer um esforço consciente para ficar chocados o suficiente.

Referências

1. *David G. Meyers, Exploring Social Psychology (New York: McGraw-Hill, 1994), 15–19.*
2. *Paul F. Lazarsfeld, "The American Solidier—An Expository Review," Public Opinion Quarterly 13, no. 3 (1949): 377–404.*
3. *Daphna Baratz, How Justified Is the "Obvious" Reaction? (Stanford University, 1983).*

D — Respostas misteriosas

30 — Explicações falsas

Era uma vez um instrutor que ensinava física aos alunos. Um dia, ele os chamou para a sala de aula e mostrou-lhes uma grande placa quadrada de metal ao lado de um radiador quente. Cada aluno colocou a mão na placa e sentiu que o lado próximo ao radiador estava frio, enquanto o lado distante estava quente. O instrutor perguntou: "Por que você acha que isso acontece?" Alguns alunos sugeriram que era devido à convecção das correntes de ar, enquanto outros pensaram que poderia haver metais diferentes na placa. Eles propuseram muitas explicações criativas, mas nenhum deles se rendeu à ideia de dizer "Eu não sei" ou "Isso parece impossível".

A resposta surpreendente foi que o instrutor havia girado a placa antes de os alunos entrarem na sala.[1]

Agora, considere o aluno que gagueja freneticamente e responde: "Eh, talvez seja devido à condução de calor?" Pergunto: "Será que essa crença é apropriada?" As palavras são facilmente pronunciadas e enfáticas, mas elas controlam realmente a antecipação?

Refletindo sobre a pequena frase "por causa de", que vem antes de "condução de calor", podemos pensar em outras explicações que poderíamos colocar após ela. Por exemplo, poderíamos dizer "por causa do flogisto" ou "por causa da magia".

Você pode argumentar que a frase "por causa da magia" não é uma explicação científica. Na verdade, as frases "por causa da condução de calor" e "por causa da magia" pertencem a gêneros literários diferentes. "Condução de calor" é algo que Spock poderia dizer em Star Trek, enquanto "magia" seria algo dito por Giles em Buffy, a Caça-Vampiros[11].

No entanto, como bayesianos, não nos importamos com os gêneros literários. Para nós, a substância de um modelo é o controle que ele exerce sobre a antecipação. Se você diz "condução de calor", que experiência isso o leva a antecipar? Em circunstâncias normais, isso leva você a prever que, se você colocar a mão na lateral da placa próxima ao radiador, esse lado ficará mais quente do que o lado oposto. Se "por causa da condução de calor" também pode explicar o resfriamento do lado adjacente ao radiador, então pode explicar praticamente qualquer coisa.

E, como todos nós sabemos a essa altura (espero), se você for igualmente bom em explicar qualquer resultado, não terá conhecimento. "Por causa da condução de calor", usado dessa maneira, é uma hipótese disfarçada de entropia máxima. É uma antecipação isomórfica para dizer "mágica". Parece uma explicação, mas não é.

Suponha que, em vez de adivinhar, medimos o calor da placa de metal em vários pontos e em vários momentos. Vendo uma placa de metal próxima ao radiador, normalmente esperaríamos que as temperaturas pontuais satisfizessem um equilíbrio da equação de difusão em relação às condições de contorno impostas pelo ambiente. Você pode não saber a temperatura exata do primeiro ponto medido, mas após medir os primeiros pontos — não sou físico o suficiente para saber quantos seriam necessários — você pode adivinhar o resto.

---

11 NT: Seriado norte-americano popular na década de 1990. Título original em inglês: *Buffy, the vampire slayer.*

Um verdadeiro mestre na arte de usar números para restringir a antecipação de fenômenos materiais — um físico — faria algumas medições e diria: “Esta placa estava em equilíbrio com o ambiente há dois minutos e meio, virou e está agora se aproximando do equilíbrio novamente”.

O erro mais profundo dos alunos não é simplesmente que eles falharam em restringir a antecipação. O erro mais profundo deles é que eles pensaram estarem fazendo física. Eles diziam a frase “por causa de”, seguida pelo tipo de palavras que Spock poderia dizer em Jornada nas Estrelas, e achavam que assim entravam no magistério da ciência.

Não é assim. Eles simplesmente mudaram sua magia de um gênero literário para outro.

Referências

1. *Search for “heat conduction.” Taken from Joachim Verhagen,http://web.archive.org/web/20060424082937/http://www.nvon.nl/scheik/best/diversen/scijokes/scijokes.txt, archived version, October 27, 2001.*

31 — Adivinhando a senha do professor

Quando era jovem, eu li livros populares de física, como o *QED*[12] de Richard Feynman: *The Strange Theory of Light and Matter* (A estranha teoria da luz e da matéria). Eu sabia que a luz era onda, o som era onda e a matéria era onda. Fiquei orgulhoso do meu conhecimento científico quando tinha nove anos.

Ao ficar mais velho e começar a ler as Feynman *Lectures on Physics* (Palestras sobre física de Feynman), encontrei uma joia chamada "a equação da onda". Pude acompanhar a derivação da equação, mas, olhando para trás, não pude ver sua verdadeira natureza à primeira vista. Então, pensei na equação da onda por três dias, intermitentemente, até que vi que era embaraçosamente óbvio. E quando finalmente entendi, percebi que durante todo o tempo em que aceitei a garantia honesta dos físicos de que a luz era onda, o som era onda e a matéria era onda, eu não tinha a mais vaga ideia do que a palavra "onda" significava para um físico.

Existe uma tendência instintiva em pensar que, se um físico disser "a luz é feita de ondas", e o professor perguntar "Do que é feita a luz?", e o aluno responder "Ondas!", então o aluno está fazendo uma afirmação verdadeira. Isso é justo, certo? Aceitamos "ondas" como a resposta correta do físico; não seria injusto rejeitá-lo do aluno? Certamente, a resposta "Ondas!" é verdadeira ou falsa, certo?

Este é apenas mais um mau hábito a ser desaprendido na escola. As palavras não têm definições intrínsecas. Se eu ouvir as sílabas "cas-tor" e pensar em um grande roedor, isso é um fato sobre meu próprio estado de espírito, não um fato sobre as sílabas "castor". A sequência de sílabas "feita de ondas" (ou "devido à condução de calor") não é uma hipótese, é um padrão de vibrações que viaja pelo ar ou é impresso em papel. Pode-se associar a uma hipótese na cabeça de alguém, mas não é, por si só, certo ou errado. Mas na escola, o professor dá a você uma estrela dourada por dizer "feita de ondas", que deve ser a resposta correta, porque o professor ouviu um físico emitindo as mesmas vibrações sonoras. Como o comportamento verbal (falado ou escrito) é o que recebe a estrela dourada, os alunos começam a pensar que o comportamento verbal tem um valor de verdade. Afinal, ou a luz é feita de ondas, ou não, certo?

E isso leva a um hábito ainda pior. Suponha que o professor apresente a você um problema confuso envolvendo uma placa de metal próxima a um radiador; o lado mais distante parece mais quente do que o lado próximo ao radiador. A professora pergunta: "Por quê?" Se você responder "não sei", não terá chance de ganhar uma estrela de ouro — nem contará como participação na aula. Mas durante este semestre, este professor usou frases como "por causa da convecção de calor", "por causa da condução de calor" e "por causa do calor radiante". Um deles é provavelmente a resposta correta que o professor deseja. Você pergunta: "Eh, talvez por causa da condução de calor?"

Essa não é uma hipótese sobre a placa de metal. Na verdade, não é nem mesmo uma suposição adequada. É uma tentativa de adivinhar a resposta que o professor quer.

12 NT: *Quantum Electric Dynamics* — Eletrodinâmica quântica.

Mesmo que você visualize os símbolos da equação de difusão (a matemática que governa a condução de calor), isso não significa que você formou uma hipótese sobre a placa de metal. Isso não é uma escola tradicional; não estamos testando sua memória para ver se você pode escrever a equação de difusão. Isso é Bayescraft[13]; estamos marcando suas expectativas de experiência. Se você usar a equação de difusão, medindo alguns pontos com um termômetro e tentando prever o que o termômetro dirá na próxima medição, então ela está definitivamente ligada à experiência. Mesmo que o aluno apenas visualize algo fluindo e, portanto, segure um fósforo perto do lado mais frio do prato para tentar medir para onde vai o calor, essa imagem mental de fluidez se conecta à experiência e controla as expectativas.

Se você não estiver usando a equação de difusão — colocando números e obtendo resultados que controlam suas expectativas de experiências particulares, então a conexão entre mapa e território é cortada como se fosse por uma faca. O que resta não é uma crença, mas um comportamento verbal.

No sistema escolar, tudo gira em torno do comportamento verbal, seja escrito no papel ou falado em voz alta. O comportamento verbal dá a você uma estrela de ouro ou uma nota de reprovação. Desaprender esse mau hábito envolve tornar-se consciente da diferença entre uma explicação e uma resposta desejada.

Isso parece muito difícil? Quando você se depara com uma placa de metal confusa, não poderia "condução de calor?" ser o primeiro passo para encontrar a resposta? Talvez, mas somente se você não cair na armadilha de pensar que está procurando uma senha. E se não houver um professor para lhe dizer que você falhou? Então, você pode pensar que "Luz é feita de wakalixes" é uma boa explicação, que "wakalixes" é a senha correta. Isso aconteceu comigo quando eu tinha nove anos — não porque eu fosse estúpido, mas porque é o que acontece por padrão. É assim que os seres humanos pensam, a menos que sejam treinados para não cair na armadilha. A humanidade ficou presa em buracos como esse por milhares de anos.

Talvez, se ensinarmos aos alunos que as palavras não contam, apenas os controladores de antecipação, eles não ficarão presos em "condução de calor"? Não. Talvez "convecção de calor"? Também não é isso? Então, pensar que a frase "condução de calor" levará a um caminho genuinamente útil, como:

- "Condução de calor?"
- Mas isso é apenas uma frase — o que significa?
- A equação de difusão?
- Mas esses são apenas símbolos — como aplicá-los?
- O que a aplicação da equação de difusão me leva a antecipar?
- Certamente não me leva a antecipar que o lado de uma placa de metal mais distante de um radiador pareceria mais quente.
- Percebo que estou confuso. Talvez o lado mais próximo pareça mais frio porque é feito de material mais isolante e transfere menos calor para minha mão. Tentarei medir a temperatura…
- Ok, não foi isso. Posso tentar verificar se a equação de difusão é verdadeira para esta placa de metal? O calor está fluindo normalmente ou há algo mais acontecendo?
- Eu poderia apontar um fósforo para o prato e tentar medir como o calor se espalha ao longo do tempo…

Se não formos rigorosos sobre "Eh, talvez por causa da condução de calor?" ser uma explicação falsa, o aluno muito provavelmente ficará preso em alguma senha "wakalixes". Isso acontece por padrão: aconteceu com toda a espécie humana por milhares de anos.

13 NT: O termo não tem tradução oficial para o português brasileiro, mas pode ser traduzida como Arte Bayesiana.

32 — Ciência como vestimenta

A prévia do filme dos *X-Men* apresenta uma narração que diz: “Em cada ser humano… existe o código genético… para mutação”. Aparentemente, é possível adquirir vários tipos de habilidades legais por meio da mutação. Por exemplo, a mutante Tempestade tem a habilidade de lançar raios.

Peço ao leitor que considere a maquinaria biológica necessária para gerar eletricidade, as adaptações biológicas necessárias para evitar danos causados pela eletricidade e os circuitos cognitivos necessários para o controle preciso dos relâmpagos. Se observássemos um organismo adquirir essas habilidades em uma geração como resultado de uma mutação, isso desacreditaria completamente o modelo neodarwiniano de seleção natural. Seria pior do que encontrar fósseis de coelhos no pré-cambriano. Se a teoria da evolução pudesse realmente explicar a existência de Tempestade, ela conseguiria explicar qualquer coisa, e todos sabemos o que isso implicaria.

Os quadrinhos dos X-Men usam termos como “evolução”, “mutação” e “código genético” apenas para se enquadrar no que eles consideram o gênero literário da ciência. A parte que me preocupa é saber quantas pessoas, principalmente na mídia, entendem a ciência apenas como um gênero literário.

Encontro pessoas que definitivamente acreditam na evolução e zombam da loucura dos criacionistas. No entanto, elas não têm ideia do que a teoria da biologia evolutiva permite e proíbe. Elas falam sobre “o próximo passo na evolução da humanidade”, como se a seleção natural tivesse um plano preconcebido. Ou pior ainda, elas falam sobre algo completamente fora do domínio da biologia evolutiva, como o aprimoramento do design de chips de computador, divisão de corporações ou humanos se transferindo para computadores, e chamam isso de “evolução”. Se a biologia evolutiva pudesse abranger isso, poderia abranger qualquer coisa.

Provavelmente, a maioria das pessoas que acredita na evolução usa a frase “por causa da evolução” porque quer fazer parte da multidão científica — crença como traje científico, como vestir um jaleco. Se a multidão científica, em vez disso, usasse a frase “por causa do design inteligente”, eles a usariam com a mesma alegria — não faria diferença para seus controladores de antecipação. Dizer “por causa da evolução” em vez de “por causa do design inteligente” não proíbe, para eles, a mutante Tempestade. Seu único propósito, para eles, é se identificar com uma tribo.

Encontro pessoas que estão bastante dispostas a considerar a ideia de uma inteligência artificial menos inteligente que a humana, ou até mesmo uma inteligência artificial ligeiramente mais inteligente que a humana. No entanto, se apresento a ideia de uma inteligência artificial fortemente sobre-humana, eles de repente decidem ser “pseudociência”. Não é porque eles pensam que têm uma teoria da inteligência que lhes permite calcular um limite teórico superior no poder de um processo de otimização. Em vez disso, eles associam uma IA fortemente sobre-humana ao gênero literário da literatura apocalíptica, enquanto uma IA que dirige uma pequena corporação se associa ao gênero literário da revista Wired. Eles não estão falando de um modelo de cognição. Eles não percebem que precisam de um modelo. Eles não percebem que a ciência é sobre modelos. Suas críticas devastadoras consistem puramente em comparações com a literatura apocalíptica, em vez de, por exemplo, leis conhecidas que proíbem tal resultado. Eles entendem a ciência apenas como um gênero literário ou um grupo ao qual pertencer. O traje não parece para eles um jaleco; este não é o time de futebol que eles estão torcendo.

Existe alguma ideia científica na qual você acredita, embora não a utilize profissionalmente? É melhor você se perguntar quais experiências futuras essa crença proíbe de acontecer com você. Essa é a soma do que você assimilou e fez parte de si mesmo. Qualquer outra coisa, provavelmente, são senhas ou vestimentas.

33 — Causalidade falsa

O flogisto foi a resposta do século XVIII para o Fogo Elemental dos alquimistas gregos. Acenda a madeira e deixe queimar. O que é o material de "fogo" alaranjado? Por que a madeira se transforma em cinzas? Para ambas as perguntas, os químicos do século XVIII responderam: "flogisto".

E foi isso, você vê, essa foi a resposta deles: "Flogisto." Flogisto escapou de substâncias em chamas como fogo visível. À medida que o flogisto escapava, as substâncias que queimavam perdiam o flogisto e assim se tornavam cinzas, o "verdadeiro material". As chamas em recipientes fechados se apagaram porque o ar ficou saturado de flogisto e, portanto, não aguentou mais. O carvão vegetal deixava poucos resíduos ao queimar porque era quase flogisto puro.

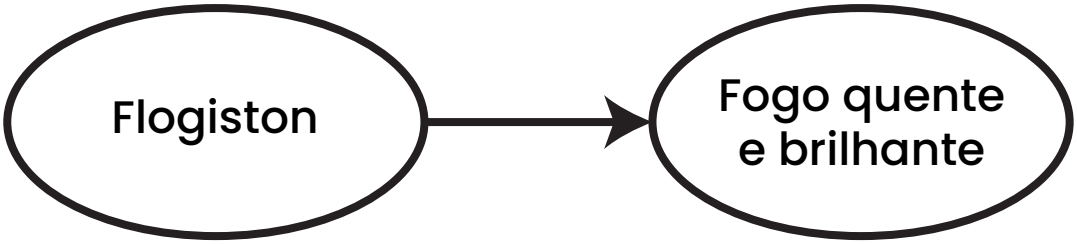
Claro, não se usava a teoria do flogisto para prever o resultado de uma transformação química. Você olhava primeiro para o resultado e depois usava a teoria do flogisto para explicá-lo. Não é que os teóricos do flogisto previram que uma chama se extinguiria em um recipiente fechado; em vez disso, acendiam uma chama em um recipiente, observavam-na apagar e então diziam: "O ar deve ter ficado saturado de flogisto". Nem se poderia usar a teoria do flogisto para dizer o que não deveria ver; poderia explicar tudo.

Essa foi uma era anterior à ciência. Por muito tempo, ninguém percebeu haver um problema. Explicações falsas não parecem falsas. É isso que as torna perigosas.

Pesquisas modernas sugerem que os humanos pensam sobre causa e efeito usando algo como os grafos acíclicos direcionados (DAGs) das redes de Bayes. Porque choveu, a calçada está molhada; porque a calçada está molhada, está escorregadia:

A partir disso, podemos inferir — ou, em uma rede de Bayes, calcular rigorosamente em probabilidades — que, quando a calçada está escorregadia, provavelmente choveu; mas se já sabemos que a calçada está molhada, saber que a calçada está escorregadia não nos diz mais nada sobre se choveu.

Por que o fogo é quente e brilhante quando queima?



A explicação do fogo ser quente e brilhante quando queima pode parecer válida, mas a mente humana não detecta automaticamente quando uma causa tem um efeito irrestrito. Além disso, o viés retrospectivo pode levar a crer que a causa restringe o efeito, quando ela foi meramente ajustada ao efeito.

Nesse sentido, nossa compreensão moderna do raciocínio probabilístico sobre causalidade pode precisamente descrever o que os teóricos do flogisto estavam fazendo de errado. Uma das principais inspirações para as redes bayesianas foi perceber o problema da contagem dupla de evidências se a inferência ressoar entre um efeito e uma causa. Por exemplo, se recebemos informações não confiáveis de que a calçada está molhada, isso nos faz pensar que é mais provável que esteja chovendo. No entanto, se está chovendo, a calçada provavelmente está molhada e escorregadia, o que aumenta ainda mais a probabilidade de que esteja chovendo.

Judea Pearl usa a metáfora de um algoritmo para contar soldados em uma fila. Suponha que você esteja na fila e veja dois soldados ao seu lado, um na frente e outro atrás. São três soldados, incluindo você. Então, você pergunta ao soldado atrás de você: "Quantos soldados você vê?" Eles olham em volta e dizem: "Três". Agora são seis soldados no total. Obviamente, isso não é a maneira mais inteligente de contar. Uma maneira mais eficiente é perguntar ao soldado à sua frente: "Quantos soldados há à sua frente?" e ao soldado atrás, "Quantos soldados há atrás de você?" A pergunta "Quantos soldados há à frente?" pode ser transmitida como uma mensagem sem confusão. Se estiver na frente da fila, passo a mensagem "1 soldado à frente" para mim mesmo. A pessoa diretamente atrás de mim recebe a mensagem "1 soldado à frente" e passa a mensagem "2 soldados à frente" para o soldado atrás dela. Ao mesmo tempo, cada soldado também recebe a mensagem "N soldados atrás" do soldado atrás deles e a transmite como "N + 1 soldados atrás" para o soldado à sua frente. Quantos soldados no total? Adicione os dois números que você recebeu, mais um para você: esse é o número total de soldados em linha.

A ideia principal é que cada soldado deve rastrear separadamente as duas mensagens, a mensagem de encaminhamento e a mensagem de retorno, e adicioná-las apenas no final. Você nunca adiciona soldados da mensagem de retorno que recebe à mensagem de encaminhamento que envia. Na verdade, o número total de soldados nunca é transmitido como uma mensagem — ninguém nunca o diz em voz alta.

Um princípio análogo opera no raciocínio probabilístico rigoroso sobre a causalidade. Se você aprender algo sobre se está chovendo, de alguma fonte que não seja a observação de que a calçada está molhada, isso enviará uma mensagem de encaminhamento para o próximo nó e aumentará nossa expectativa de que a calçada esteja molhada. Se você observar que a calçada está molhada, isso envia uma mensagem inversa à nossa crença de que está chovendo, e essa mensagem se propaga para todos os nós vizinhos, exceto o nó [Calçada Molhada]. Contamos cada evidência exatamente uma vez; nenhuma mensagem de atualização "salta" para frente e para trás. O algoritmo exato pode ser encontrado no clássico *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Raciocínio probabilístico em sistemas inteligentes: redes de inferência plausível) de Judea Pearl.

Então, o que deu errado na teoria do flogisto? Quando observamos que o fogo é quente, o nodo [Fogo] pode enviar evidências retrógradas para o nodo [Flogisto], levando-nos a atualizar nossas crenças sobre o flogisto. Mas, se assim for, não podemos contar isso como uma previsão bem-sucedida da teoria do flogisto. A mensagem deve ir em apenas uma direção e não retornar.

Infelizmente, os seres humanos não usam um algoritmo rigoroso para atualizar as redes de crenças. Aprendemos sobre os nodos-pais observando os filhos e prevemos os nodos-filhos a partir de crenças sobre os pais. No entanto, não mantemos livros rigorosamente separados para a mensagem de retorno e a mensagem de avanço. Apenas lembramos que o flogisto é quente, o que faz com que o fogo seja quente. Portanto, parece que a teoria do flogisto prevê o calor do fogo. Ou, pior ainda, parece que o flogisto faz o fogo ficar quente.

Até você perceber que nenhuma previsão antecipada está sendo feita, o nó causal não restritivo não é rotulado como "falso". É representado da mesma forma que qualquer outro nó em sua rede de crenças. Parece um fato, assim como todos os outros fatos que você conhece: o flogisto torna o fogo quente.

Uma IA projetada corretamente notaria o problema instantaneamente. Isso nem exigiria um código de propósito especial, apenas a contabilidade correta da rede de crenças. (Infelizmente, nós, humanos, não podemos reescrever nosso próprio código, como uma IA bem projetada poderia.)

Falar em “viés retrospectivo” é apenas uma maneira não técnica de dizer que os humanos não separam rigorosamente as mensagens diretas e inversas, permitindo que as mensagens diretas sejam contaminadas pelas mensagens inversas.

Aqueles que seguiram por muito tempo o caminho do flogisto não estavam tentando ser tolos. Nenhum cientista quer ficar deliberadamente preso em um beco sem saída. Existe alguma explicação falsa em sua mente? Se houver, garanto que ela não é rotulada como “explicação falsa”. Portanto, pesquisar seus pensamentos sobre a palavra-chave “falsa” não revelará nada.

Graças ao viés retrospectivo, não é suficiente verificar o quão bem sua teoria “prevê” fatos que você já conhece. Você tem que prever para amanhã, não para ontem. Essa é a única maneira de garantir que uma mente humana confusa envie uma mensagem direta pura.

Referências

1. *Judea Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference (San Mateo, CA: Morgan Kaufmann, 1988).*

34 — Sinais de parada semânticos

E a criança perguntou:

P: De onde veio essa pedra?

R: Eu a arranquei da grande pedra no centro da aldeia. P: De onde veio essa pedra?

R: Provavelmente, ela rolou da enorme montanha que se ergue acima de nossa aldeia.

P: De onde veio a montanha?

R: Do mesmo lugar de toda pedra — dos ossos de Ymir, o gigante primordial.

P: De onde veio o gigante primordial, Ymir? R: Do grande abismo, Ginnungagap.

P: De onde veio o grande abismo, Ginnungagap? R: Nunca faça essa pergunta.

Considere o aparente paradoxo da Primeira Causa. A ciência rastreou os eventos até o Big Bang, mas por que o Big Bang aconteceu? É fácil dizer que o zero do tempo começa no Big Bang — que não há nada antes do Big Bang no fluxo normal de minutos e horas. Mas dizer isso pressupõe nossa lei física, que parece altamente estruturada e clama por uma explicação. De onde vieram as leis físicas? Você poderia dizer que somos todos uma simulação de computador, mas a simulação de computador é baseada nas leis da física de algum outro mundo — de onde vieram essas leis da física?

Nesse ponto, algumas pessoas dizem: “Deus!”

O que poderia fazer alguém, mesmo uma pessoa altamente religiosa, pensar que isso ajudou a responder ao paradoxo da Primeira Causa? Por que você não perguntaria automaticamente: “De onde veio Deus?” Dizer “Deus não tem causa” ou “Deus criou a si mesmo” nos deixa exatamente na mesma posição de “O tempo começou com o Big Bang”. Ainda assim, perguntamos por que todo o metassistema existe em primeiro lugar ou por que alguns eventos, mas não outros, podem ser incausados.

Meu propósito aqui não é discutir o aparente paradoxo da Primeira Causa, mas perguntar por que alguém pensaria que “Deus!” poderia resolver o paradoxo. Dizer “Deus!” é uma forma de pertencer a uma tribo que dá às pessoas um motivo para dizer isso com a maior frequência possível — algumas pessoas até dizem isso para perguntas como “Por que esse furacão atingiu Nova Orleans?” Mesmo assim, você esperaria que as pessoas percebessem que, no quebra-cabeça específico da Primeira Causa, dizer “Deus!” não ajuda. Isso não faz o paradoxo parecer menos paradoxal, mesmo que seja verdadeiro. Como alguém poderia não perceber isso?

Jonathan Wallace sugeriu que “Deus!” funciona como um sinal semântico de parada — não é uma afirmação proposicional, mas um sinal de trânsito cognitivo: não pense além desse ponto. Dizer “Deus!” não resolve tanto o paradoxo, mas coloca um sinal de trânsito cognitivo para interromper a continuação óbvia da cadeia de perguntas e respostas.

Claro, você nunca faria isso, sendo um bom ateu, certo? Mas “Deus!” não é o único sinal de parada semântico, apenas o primeiro exemplo óbvio.

As tecnologias transumanas, tais como a nanotecnologia molecular, a biotecnologia avançada, a tecnologia genética e a inteligência artificial, apresentam desafios políticos difíceis. Que papel, se houver, o governo deveria assumir na supervisão da escolha dos genes pelos pais para seus filhos? Os pais deveriam ter permissão para escolher genes relacionados à esquizofrenia para seus filhos? Se melhorar a inteligência de uma criança for caro, os governos deveriam ajudar a garantir o acesso para evitar o surgimento de uma elite cognitiva? Embora se possam propor várias instituições para responder a essas questões políticas, como instituições de caridade privadas que possam fornecer ajuda financeira para o aprimoramento da inteligência, a próxima pergunta óbvia é: "Essa instituição será eficaz?" Se contarmos com ações judiciais de responsabilidade do produto para impedir que as corporações construam nanotecnologia prejudicial, isso funcionará realmente?

Conheço alguém cuja resposta para cada uma dessas perguntas é "democracia liberal!" Essa é a resposta dele. Se você fizer a pergunta óbvia: "Quão bem as democracias liberais se saíram historicamente em problemas tão complicados?" ou "E se a democracia liberal fizer algo estúpido?", então você é rotulado como autocrata, libertário ou uma pessoa péssima. Ninguém pode questionar a democracia.

Uma vez chamei esse tipo de pensamento de "o direito divino da democracia", mas é mais preciso dizer que "Democracia!" funcionou para ele como um sinal semântico de parada. Se alguém tivesse dito a ele "Entregue para a corporação Coca-Cola!", ele teria feito as próximas perguntas óbvias: Por quê? O que a corporação Coca-Cola fará a respeito? Por que devemos confiar neles? Eles se saíram bem no passado em problemas igualmente complicados?

Ou suponha que alguém diga: "Mexicanos-americanos estão planejando remover todo o oxigênio da atmosfera da Terra". Você provavelmente perguntaria: "Por que eles fariam isso? Os mexicanos-americanos também não precisam respirar? Os mexicanos-americanos estão todos conspirando juntos?" Se você não fizer essas próximas perguntas óbvias, quando alguém disser: "Corporações estão planejando remover o oxigênio da Terra", então "Corporações!" funciona para você como um sinal de parada semântico.

Tenha cuidado para não criar um novo argumento genérico contra coisas de que você não gosta, como "Ah, é apenas um sinal de parada!" Nenhuma palavra é um sinal de parada em si mesma. A questão é se uma palavra tem esse efeito sobre uma pessoa em particular. Ter emoções fortes sobre algo não o qualifica como um sinal de parada. Posso não gostar de terroristas ou ter medo da propriedade privada, mas isso não significa que "terroristas!" ou "capitalismo!" sejam sinais de trânsito cognitivos para mim. Embora a palavra "inteligência" já tenha tido esse efeito em mim, não tem mais. O que distingue um sinal de parada semântica é a falha em considerar a próxima pergunta óbvia.

35 — Respostas misteriosas para perguntas misteriosas

Imagine olhar para a sua mão e não saber nada sobre células, bioquímica ou DNA. Você aprendeu um pouco de anatomia através da dissecação, então sabe que a sua mão contém músculos, mas não sabe por que eles se movem em vez de ficarem parados como argila. Sua mão é uma combinação de várias coisas e, por algum motivo, se move sob a sua direção. Isso não é mágico?

> "O corpo animal não funciona como um motor termodinâmico. A consciência ensina a cada indivíduo que ele está, até certo ponto, sujeito à direção de sua vontade. Parece, portanto, que criaturas animadas podem aplicar, imediatamente, a certas partículas móveis de matéria dentro de seus corpos, forças pelas quais os movimentos dessas partículas são direcionados para produzir efeitos mecânicos derivados. A influência da vida animal ou vegetal sobre a matéria está infinitamente além do alcance de qualquer investigação científica até agora iniciada. Seu poder de dirigir os movimentos das partículas em movimento, no demonstrado milagre diário de nosso livre-arbítrio humano e no crescimento de geração após geração de plantas a partir de uma única semente, são infinitamente diferentes de qualquer resultado possível da coincidência fortuita de átomos. Os biólogos modernos estavam chegando mais uma vez à aceitação de algo, e esse era um princípio vital[14]."
>
> — Lorde Kelvin [1]

Essa era a teoria do vitalismo, que acreditava que a misteriosa diferença entre matéria viva e inanimada era explicada por um elã vital ou *vis vitalis*. O elã vital infundia a matéria viva e fazia com que ela se movesse conforme a direção consciente. Ele também participava de transformações químicas que nenhuma mera partícula não viva poderia sofrer — a síntese posterior de Wöhler de ureia, um componente da urina, foi um grande golpe para a teoria vitalista, ao mostrar que a mera química poderia duplicar um produto da biologia.

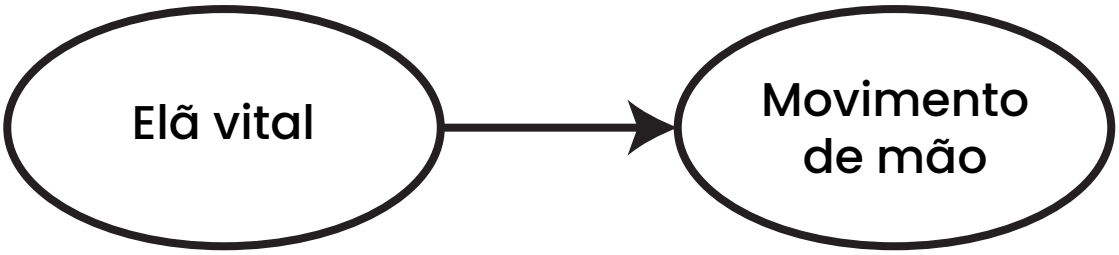
Chamar de "elã vital" uma explicação, mesmo uma explicação falsa como flogisto, provavelmente está dando muito crédito a isso. A teoria funcionava principalmente como uma rolha para a curiosidade. Quando você perguntava "Por quê?", a resposta era "elã vital!".

Dizer "elã vital!" parece explicar por que a sua mão se move.

Mas, na verdade, você tem apenas um pequeno diagrama causal em sua cabeça que diz:

---

14 NT: Tradução livre do texto original em inglês. *The animal body does not act as a thermodynamic engine consciousness teaches every individual that they are, to some extent, subject to the direction of his will. It appears therefore that animated creatures have the power of immediately applying to certain moving particles of matter within their bodies, forces by which the motions of these particles are directed to produce derived mechanical effects… The influence of animal or vegetable life on matter is infinitely beyond the range of any scientific inquiry hitherto entered on. Its power of directing the motions of moving particles, in the demonstrated daily miracle of our human free-will, and in the growth of generation after generation of plants from a single seed, are infinitely different from any possible result of the fortuitous concurrence of atoms… Modern biologists were coming once more to the acceptance of something and that was a vital principle.*



Mas, na verdade, você não sabe nada que não soubesse antes. Você não sabe, por exemplo, se a sua mão irá gerar calor ou absorver calor, a menos que já tenha observado isso; caso contrário, não poderá prever com antecedência. Sua curiosidade parece ter sido saciada, mas não foi alimentada. Uma vez que você pode dizer "Por quê? Elã vital!" para qualquer observação possível, isso é igualmente bom para explicar todos os resultados, o que é uma hipótese disfarçada de máxima entropia, entre outras.

No entanto, a maior lição reside na reverência dos vitalistas pelo elã vital e sua ânsia de declará-lo um mistério além de toda ciência. Quando encontraram o grande dragão Desconhecido, os vitalistas não desembainharam suas espadas para a batalha, mas curvaram seus pescoços em submissão. Eles se orgulhavam da sua ignorância, faziam da biologia um mistério sagrado e, assim, relutavam em prescindir de sua ignorância quando as evidências batiam à porta.

O Segredo da Vida estava infinitamente além do alcance da ciência! Não apenas um pouco além, mas infinitamente além! Lord Kelvin certamente teve uma grande emoção ao se dar conta de que não sabia de algo.

Mas a ignorância existe no mapa, não no território. Se sou ignorante sobre um fenômeno, isso é um fato sobre meu próprio estado de espírito, não um fato sobre o fenômeno em si. Um fenômeno pode parecer misterioso para uma pessoa em particular. Não há fenômenos que sejam misteriosos por si mesmos. Adorar um fenômeno porque parece tão maravilhosamente misterioso é adorar a própria ignorância.

O vitalismo compartilhava com o flogisto o erro de encapsular o mistério como uma substância. O fogo era misterioso, e a teoria do flogisto encapsulava o mistério em uma substância misteriosa chamada "flogisto". A vida era um mistério sagrado, e o vitalismo encapsulava o mistério sagrado em uma substância misteriosa chamada "elã vital". Nenhuma das respostas ajudou a concentrar a densidade de probabilidade do modelo — tornando alguns resultados mais fáceis de explicar do que outros. A "explicação" apenas envolvia a questão como uma pequena bola preta dura e opaca.

Em uma comédia escrita por Molière, um médico explica o poder de um soporífero dizendo que ele contém uma "potência dormitiva". O mesmo princípio se aplica aqui. É uma falha da psicologia humana que, diante de um fenômeno misterioso, postulemos mais facilmente substâncias inerentes misteriosas do que processos subjacentes complexos.

Mas a falha mais profunda é acreditar que uma resposta possa ser misteriosa. Se um fenômeno parece misterioso, isso é um fato sobre nosso estado de conhecimento e não sobre o fenômeno em si. Os vitalistas viram uma lacuna misteriosa em seu conhecimento e postularam algo misterioso para preencher a lacuna. Ao fazer isso, confundiram o mapa com o território. Toda confusão e perplexidade existem na mente e não em substâncias encapsuladas.

Essa é a explicação definitiva e totalmente geral do motivo pelo qual, repetidas vezes na história da humanidade, as pessoas ficam chocadas ao descobrir que uma pergunta incrivelmente misteriosa tem uma resposta não misteriosa. O mistério é uma propriedade das perguntas e não das respostas.

Portanto, eu chamo de respostas misteriosas para perguntas misteriosas teorias como o vitalismo.

Estes são os sinais de respostas misteriosas para perguntas misteriosas:

- Em primeiro lugar, a explicação funciona como um limitador da curiosidade, em vez de um controlador da antecipação.
- Em segundo lugar, a hipótese não tem partes móveis — o modelo não é um mecanismo complexo específico, mas uma substância ou força inexpressivamente sólida. Pode-se dizer que a substância misteriosa ou força misteriosa está aqui ou ali, para causar isto ou aquilo; mas a razão pela qual a força misteriosa se comporta assim está envolta em uma unidade vazia.
- Em terceiro lugar, aqueles que oferecem a explicação, estimam a sua ignorância; eles falam com orgulho de como o fenômeno derrota a ciência comum ou é diferente de fenômenos meramente mundanos.
- Em quarto lugar, mesmo depois que a resposta é dada, o fenômeno ainda é um mistério e possui a mesma qualidade de maravilhosa inexplicabilidade que tinha no início.

Referências

1. *Silvanus Phillips Thompson, The Life of Lord Kelvin (American Mathematical Society, 2005).*

36 — A futilidade da emergência

As falhas do flogisto e do vitalismo são consideradas retrocessos históricos. Posso me arriscar a citar uma teoria atual que considero igualmente falha?

Refiro-me à emergência ou fenômenos emergentes, que geralmente se referem ao estudo de sistemas em que comportamentos complexos surgem da interação de muitos elementos simples. (Conforme a Wikipedia, "é a maneira como sistemas e padrões complexos surgem de uma multiplicidade de interações relativamente simples".) Se tomarmos essa descrição literalmente, ela se aplicaria a todos os fenômenos acima do nível de quarks individuais em nosso universo, o que é problemático. Imagine apontar para um colapso do mercado financeiro e dizer "Não é um quark!" Isso explica alguma coisa? Claro que não. Então, assim como não faria sentido usar "quark" para explicar o mercado financeiro, não faz sentido usar "emergência" para explicar sistemas complexos.

É o termo "emergência" que me incomoda, não o verbo "emergir". Não há nada de errado em dizer que "X emerge de Y", onde Y é um modelo detalhado com partes internas móveis. "Surge de" é outra frase válida que significa o mesmo. Por exemplo, a gravidade surge da curvatura do espaço-tempo, conforme o modelo matemático específico da Relatividade Geral, enquanto a química surge das interações entre átomos, segundo o modelo específico da eletrodinâmica quântica.

Agora, suponha que eu queira explicar a gravidade dizendo que ela "surgiu" ou que a química é um "fenômeno que surge". Isso não seria satisfatório, a menos que eu apresente um modelo detalhado para avaliação.

A frase "emerge de" é aceitável, assim como "surgiu de" ou "é causado por", se preceder um modelo específico que possa ser avaliado por seus próprios méritos.

No entanto, esse não é o jeito como "emergência" é comumente usada. As pessoas costumam usar "emergência" como uma explicação por si só.

Perdi a conta de quantas vezes ouvi as pessoas dizerem: "A inteligência é um fenômeno emergente!" como se isso explicasse a inteligência. Esse uso se encaixa em todos os itens da lista de verificação para uma resposta misteriosa a uma pergunta misteriosa. O que você sabe, após ter dito que a inteligência é "emergente"? Você não pode fazer novas previsões. Você não sabe nada sobre o comportamento das mentes reais que não soubesse antes. Parece que você acredita em um fato novo, mas não antecipa nenhum resultado diferente. Sua curiosidade parece saciada, mas não foi alimentada. A hipótese não tem partes móveis — não há modelo interno detalhado para manipular. Aqueles que proferem a hipótese de "emergência" confessam sua ignorância dos internos e se orgulham disso; eles contrastam a ciência da "emergência" com outras ciências meramente mundanas.

E mesmo após a resposta de "Por quê? Emergência!" é dado, o fenômeno ainda é um mistério e possui a mesma impenetrabilidade sagrada que tinha no início.

Um exercício interessante é eliminar o adjetivo “emergente” de qualquer frase em que apareça e ver se a frase diz algo diferente:

- Antes: A inteligência humana é um produto emergente do disparo de neurônios.
- Depois: A inteligência humana é um produto do disparo de neurônios.
- Antes: O comportamento da colônia de formigas é o resultado emergente das interações de muitas formigas individuais.
- Depois: O comportamento da colônia de formigas é o resultado das interações de muitas formigas individuais.
- Melhor ainda: uma colônia é feita de formigas. Podemos prever com sucesso alguns aspectos do comportamento da colônia usando modelos que incluem apenas formigas individuais, sem nenhuma variável global da colônia, mostrando que entendemos como esses comportamentos da colônia surgem do comportamento das formigas.

Outro exercício interessante é substituir a palavra “emergente” pela palavra antiga, a explicação que as pessoas tinham que usar antes da invenção da emergência:

- Antes: a vida é um fenômeno emergente.
- Depois: a vida é um fenômeno mágico.
- Antes: a inteligência humana é um produto emergente do disparo de neurônios.
- Depois: a inteligência humana é um produto mágico do disparo de neurônios.

Cada afirmação não transmite a mesma quantidade de conhecimento sobre o comportamento do fenômeno? Cada hipótese não se encaixa exatamente no mesmo conjunto de resultados?

A palavra “emergência” tornou-se tão popular quanto a palavra ‘mágica’ costumava ser. A razão é que ambas têm um apelo profundo à psicologia humana. A explicação emergente é maravilhosamente fácil e agradável de se dizer; ela cria um mistério sagrado para adorar. A emergência é popular porque é a “junk food” da curiosidade. As pessoas podem explicar qualquer coisa usando a emergência, e muitas vezes o fazem porque é maravilhoso explicar as coisas. Os humanos ainda são humanos, mesmo quando estudam ciência na faculdade. Quando encontram uma maneira de escapar das correntes da ciência estabelecida, eles se envolvem nas mesmas travessuras de seus ancestrais, vestidas com a vestimenta do gênero literário “Ciência”, mas os humanos ainda são humanos e a psicologia humana ainda é a mesma.

37 — Não diga “complexidade”

Era uma vez. . .

Esta é uma história de quando conheci Marcello, com quem trabalhei posteriormente por um ano na teoria da Inteligência Artificial (IA), mas nessa época eu ainda não o havia aceitado como meu aprendiz. Sabia que ele competia em olimpíadas nacionais de matemática e computação, o que foi suficiente para que eu lhe desse mais atenção; no entanto, ainda não sabia se ele conseguiria aprender a pensar em IA.

Eu perguntei a Marcello como ele achava que uma IA poderia descobrir como resolver um cubo mágico por si só, não por meio de programação prévia, o que é trivial, mas sim descobrindo as leis do universo de Rubik e raciocinando como explorá-las. Como uma IA inventaria para si mesma o conceito de “operador” ou “macro”, que é a chave para resolver o cubo mágico?

Em algum momento dessa discussão, Marcello disse: “Bem, acho que a IA precisa de complexidade para fazer X e complexidade para fazer Y...”

E eu disse: “Não diga ‘complexidade’ ”. Marcello disse: “Por que não?”

Eu disse: “A complexidade nunca deve ser um objetivo em si. Você pode precisar usar um algoritmo específico que adiciona alguma complexidade, mas a complexidade pela complexidade apenas torna as coisas mais difíceis.” (Estava pensando em todas as pessoas que ouvi defendendo que a Internet “acordaria” e se tornaria uma IA quando se tornasse “suficientemente complexa”.)

E Marcello disse: “Mas deve haver alguma complexidade que faz isso.”

Fechei os olhos brevemente e tentei pensar em como explicar tudo em palavras. Para mim, dizer “complexidade” simplesmente parecia o movimento errado na dança da IA. Ninguém pode pensar rápido o suficiente para deliberar, em palavras, sobre cada frase de seu fluxo de consciência; pois isso exigiria uma recursão infinita. Pensamos em palavras, mas nosso fluxo de consciência é dirigido abaixo do nível das palavras, pelos remanescentes treinados de percepções passadas e duras experiências…

Eu disse: “Você leu uma explicação técnica da explicação técnica?” “Sim”, disse Marcello.

“Tudo bem”, eu disse. “Dizer ‘complexidade’ não concentra sua massa de probabilidade.”

“Oh,” Marcello disse, “como ‘emergência’. Ah! Então… agora tenho que pensar em como X pode realmente acontecer. . .”

Foi quando pensei comigo mesmo: “Talvez **este** seja possível ensinar.” Complexidade não é um conceito sem utilidade. Existem definições matemáticas relacionadas a ele, como a complexidade de Kolmogorov e a complexidade de Vapnik-Chervonenkis. Mesmo intuitivamente, vale a pena considerar a complexidade — é preciso avaliar a complexidade de uma hipótese e decidir se ela é “muito complicada” dada a evidência de apoio, ou analisar um projeto e tentar simplificá-lo.

Os conceitos por si só não são nem úteis, nem inúteis. Apenas seu uso pode ser correto ou incorreto. Enquanto Marcello tentava realizar um passo na dança, ele também tentava explicar algo sem esforço e obter algo sem esforço. Esse é um erro extremamente comum, pelo menos na minha área. Em uma discussão geral sobre inteligência artificial, você pode observar pessoas cometendo o mesmo erro repetidamente: pulando constantemente coisas que não entendem, sem perceber que é isso que estão fazendo.

Em um piscar de olhos, podemos colocar um nó causal não controlador atrás de algo misterioso, que parece uma explicação, mas não é. O erro ocorre abaixo do nível das palavras. Ele não requer nenhuma falha de caráter especial; é simplesmente como os seres humanos pensam por padrão, como pensavam desde tempos antigos.

O que se deve evitar é pular a parte misteriosa e, em vez disso, enfrentá-la diretamente. Há muitas palavras que podem pular os mistérios, e algumas delas seriam legítimas em outros contextos, como "complexidade", por exemplo. No entanto, o erro essencial é esse salto, independentemente de qual nó causal esteja por trás dele. O salto não é um pensamento, mas um micro-pensamento. Você tem que prestar muita atenção para detectá-lo. E quando você treina para evitar pular, isso se torna uma questão de instinto, não de raciocínio verbal. Você precisa sentir quais partes do seu mapa continuam em branco e, mais importante, prestar atenção a essa sensação.

Desconfio que na academia existe uma grande pressão para esconder os problemas debaixo do tapete, a fim de apresentar um artigo que pareça completo. É mais provável que você receba elogios por um modelo aparentemente completo que inclua alguns "fenômenos emergentes", em comparação a um mapa explicitamente incompleto que indique: "Não tenho ideia de como essa parte funciona" ou "aqui ocorre um milagre". Alguma revista pode até não aceitar esse último tipo de artigo, pois quem sabe se não são as etapas desconhecidas onde tudo de interessante acontece? E sim, às vezes acontece que todas as partes não mágicas do seu mapa também não são importantes. Esse é o preço que às vezes você paga por entrar em terra incógnita e tentar resolver problemas incrementalmente. Mas isso torna ainda mais importante saber quando você ainda não terminou. Principalmente, as pessoas não se atrevem a entrar em terra incógnita, pelo medo mortal de perder tempo.

E se você estiver trabalhando em uma startup revolucionária de IA, a pressão para esconder os problemas será ainda maior. Ou você terá que admitir para si mesmo que ainda não sabe como construir a IA, e seus planos de vida desmoronarão. Mas talvez eu esteja explicando demais, já que a tendência natural dos humanos é ignorar os problemas. Se você está procurando exemplos, observe as pessoas discutindo religião, filosofia, espiritualidade ou qualquer ciência em que não tenham sido treinadas profissionalmente.

Marcello e eu desenvolvemos uma convenção em nosso trabalho de IA: sempre que encontrávamos algo que não entendíamos — o que era bastante comum — dizíamos "mágica", como em "X faz Y magicamente". Isso nos lembrava que ali havia um problema não resolvido, uma lacuna em nossa compreensão. É muito melhor dizer "mágica" do que "complexidade" ou "fenômeno emergente", pois as últimas palavras criam a ilusão de compreensão. É mais sensato dizer "mágica" e deixar um espaço reservado para si mesmo, um lembrete do trabalho que você terá que fazer mais tarde.

38 — Viés positivo: olhe para o escuro

Estou ministrando uma aula e escrevo três números no quadro negro: 2-4-6. “Eu estou pensando em uma regra”, digo, “que governa sequências de três números. A sequência 2-4-6, por acaso, obedece a essa regra. Cada um de vocês encontrará em sua mesa um monte de cartões. Escreva uma sequência de três números em um cartão e marque “Sim” se se encaixar na regra ou “Não” se não se encaixar na regra. Então você pode escrever outro conjunto de três números e perguntar novamente se ele se encaixa, e assim por diante. Quando tiver certeza de que conhece a regra, anote-a em um cartão. Você pode testar quantos conjuntos quiser.

Aqui está o registro dos palpites de um aluno:

4-6-2 Não

4-6-8 Sim

10-12-14 Sim.

Naquele momento, o aluno anotou sua suposição sobre a regra. Qual você acredita que seja a regra? Você gostaria de testar outro trio de números e, se sim, qual seria? Tire um momento para pensar antes de continuar.

O desafio acima é inspirado em um experimento clássico criado por Peter Wason, chamado de Tarefa 2-4-6. Embora os participantes que receberam a tarefa frequentemente demonstrassem grande confiança em suas suposições, apenas 21% deles acertaram com sucesso a regra real do experimentador. Desde então, as réplicas desse experimento continuaram a mostrar taxas de sucesso em torno de 20%. [1]

O estudo recebeu o nome de On the failure to eliminate hypotheses in a conceptual task (Sobre o fracasso em eliminar uma hipótese em uma tarefa conceitual). Geralmente, os participantes que tentam a tarefa 2-4-6 tentam gerar exemplos positivos em vez de negativos — eles aplicam a hipótese da regra para gerar uma instância representativa e verificar se ela é rotulada como “Sim”.

Dessa forma, alguém que propõe a hipótese “números crescentes de dois” irá testar o trio 8-10-12, receberá a resposta “Sim”, e anunciará a regra com confiança. Alguém que propõe a hipótese “X-2X-3X” irá testar o trio 3-6-9, constatará que ele se encaixa, e anunciará essa regra.

Em todos os casos, a regra atual é a mesma: os três números devem estar em ordem crescente.

No entanto, para descobrir a regra real, seria necessário gerar trigêmeos que não deveriam se encaixar, como 20-23-26, e verificar se eles são rotulados como “Não”. Isso é algo que as pessoas tendem a não fazer neste experimento. Em alguns casos, os sujeitos inventam regras muito mais complicadas do que a resposta real, depois as “testam” e as anunciam.

Este fenômeno cognitivo é comumente conhecido como “viés de confirmação”. No entanto, na minha opinião, o fenômeno de tentar testar exemplos positivos em vez de negativos deve ser distinguido do fenômeno de tentar manter a crença com a qual você começou. “Viés positivo” às vezes é usado como sinônimo de “viés de confirmação”, mas é mais apropriado para essa falha específica.

Houve um tempo em que se pensou que a teoria do flogisto poderia explicar uma chama saindo de uma caixa fechada (o ar ficou saturado com o flogisto e nada mais poderia ser liberado). No entanto, a teoria do flogisto também poderia explicar a chama não apagando. Para perceber isso, é preciso buscar exemplos negativos em vez de exemplos positivos, olhar para o zero em vez de um, o que vai contra o instinto humano e o que a experiência tem mostrado.

Pois, por instinto, nós seres humanos só vivemos em meio mundo.

Às vezes, mesmo que uma pessoa seja instruída sobre o viés positivo por dias, ela pode ignorá-lo quando necessário. O viés positivo não é uma escolha lógica ou emocional, como na tarefa 2-4-6, o qual é um exercício frio, lógico e não afetivo. A falha acontece sub-verbalmente, no nível da imaginação e das reações instintivas. Como o problema não surge de seguir uma regra deliberada que diz: "Pense apenas em exemplos positivos", ele não pode ser resolvido apenas sabendo verbalmente que "Devemos pensar em exemplos positivos e negativos". Qual exemplo vem automaticamente à sua mente? Você deve aprender, sem palavras, a fazer zig em vez de zag. Você deve aprender a recuar em direção ao zero, em vez de se afastar dele.

Escrevo há algum tempo sobre a ideia de que a força de uma hipótese reside no que ela não consegue explicar, não no que consegue — se você consegue explicar qualquer resultado igualmente bem, você não tem conhecimento algum. Por isso, para identificar uma explicação que não ajuda, não é suficiente pensar no que ela explica muito bem — você também precisa procurar resultados que ela seria incapaz de explicar, e essa é a verdadeira força da teoria.

Então eu disse tudo isso e desafiei a utilidade da "emergência" como um conceito. Após discutir sobre a utilidade do conceito de "emergência", um comentarista citou a supercondutividade e o ferromagnetismo como exemplos de emergência. No entanto, eu afirmei que a não supercondutividade e a não magnetização também são exemplos de emergência, que era justamente o problema que eu via no uso do termo. Apesar disso, não quero criticar o comentarista. Mesmo tendo lido muito sobre o "viés de confirmação", eu não percebi imediatamente a falha na tarefa 2-4-6. É uma reação subconsciente que precisa ser desaprendida e reeducada. Eu mesmo continuo trabalhando nisso.

Grande parte da habilidade de um racionalista é abaixo do nível das palavras. É difícil tentar ensinar a Arte através da linguagem. As pessoas podem concordar com você em uma frase, mas na próxima, elas podem ter uma reação subconsciente que aponta na direção oposta. Não estou reclamando disso! Na verdade, um dos principais motivos pelos quais escrevo isso é observar o que minhas palavras não conseguem transmitir.

Você está em busca de exemplos que comprovem o viés positivo neste momento ou economizando um pouco de sua pesquisa sobre quais situações podem levá-lo a não perceber tal viés? Sua atenção está voltada para a luz ou para a escuridão?

Referências

1. *Peter Cathcart Wason, "On the Failure to Eliminate Hypotheses in a Conceptual Task," Quarterly Journal of Experimental Psychology 12, no. 3 (1960): 129–140, doi:10.1080/17470216008416717.*

39 — Incerteza legal

Em *Rational Choice in an Uncertain World* (Uma escolha racional em um mundo incerto), Robyn Dawes descreve um experimento de Tversky [1] [2]:

No final da década de 1950 e início da de 1960, diversos experimentos psicológicos foram realizados onde os participantes eram solicitados a prever o resultado de eventos com um componente aleatório, mas com uma previsibilidade básica. Um exemplo típico: os participantes tinham que adivinhar se a próxima carta que o experimentador viraria seria vermelha ou azul, em um contexto onde 70% das cartas eram azuis, embora a sequência de cartas vermelhas e azuis fosse totalmente aleatória. Nessa situação, a estratégia que garante a maior taxa de sucesso é prever o evento mais frequente. Por exemplo, se 70% das cartas são azuis, prever "azul" em cada tentativa resultaria em uma taxa de acerto de 70%.

No entanto, os participantes tendiam a combinar probabilidades, ou seja, prever o evento provável com a frequência relativa com que ele ocorreu. Por exemplo, eles costumavam prever o cartão azul 70% das vezes e o cartão vermelho 30% das vezes. Essa estratégia resulta em uma taxa de sucesso de 58%, por acertarem 70% das vezes quando o cartão azul aparece (o que acontece com probabilidade de 0,70) e 30% das vezes quando o cartão vermelho aparece (com probabilidade de 0,30); (0,70 × 0,70) + (0,30 × 0,30) = 0,58.

De fato, os participantes previam o evento mais frequente com uma probabilidade ligeiramente maior do que aquela com a qual ele ocorria, mas não chegavam perto de prever sua ocorrência 100% das vezes, mesmo quando eram pagos pela precisão de suas previsões. Por exemplo, indivíduos que recebiam um níquel por cada previsão correta em mil tentativas, previram o evento mais comum 76% das vezes[15].

Não se engane pensando que essa experiência é apenas uma pequena falha nas estratégias de jogo. Ela ilustra concisamente a ideia mais importante de toda a racionalidade.

Os participantes continuam adivinhando a cor vermelha, como se pensassem que tivessem alguma maneira de prever a sequência aleatória. Sobre esse experimento, Dawes continua dizendo: "Apesar do feedback por meio de mil tentativas, os participantes não conseguem acreditar que a situação é algo que eles não podem prever".

---

15 NT: Tradução livre do texto original em inglês. *Many psychological experiments were conducted in the late 1950s and early 1960s in which subjects were asked to predict the outcome of an event that had a random component but yet had base-rate predictability—for example, subjects were asked to predict whether the next card the experimenter turned over would be red or blue in a context in which 70% of the cards were blue, but in which the sequence of red and blue cards was totally random. In such a situation, the strategy that will yield the highest proportion of success is to predict the more common event. For example, if 70% of the cards are blue, then predicting blue on every trial yields a 70% success rate. What subjects tended to do instead, however, was match probabilities—that is, predict the more probable event with the relative frequency with which it occurred. For example, subjects tended to predict 70% of the time that the blue card would occur and 30% of the time that the red card would occur. Such a strategy yields a 58% success rate, because the subjects are correct 70% of the time when the blue card occurs (which happens with probability.70) and 30% of the time when the red card occurs (which happens with probability .30); (.70 .70) + (.30.30) = .58. In fact, subjects predict the more frequent event with a slightly higher probability than that with which it occurs, but do not come close to predicting its occurrence 100% of the time, even when they are paid for the accuracy of their predictions . . . For example, subjects who were paid a nickel for each correct prediction over a thousand trials . . . predicted [the more common event] 76% of the time.*

Mas o erro deve ir além disso. Mesmo que os participantes acreditem ter formulado uma hipótese, eles não precisam necessariamente apostar nessa previsão para testá-la. Eles podem pensar: “Se essa hipótese estiver correta, a próxima carta será vermelha” e, em seguida, apostar em azul. Poderiam escolher azul em todas as tentativas, acumulando o máximo de fichas possível, enquanto mentalmente registram suas suposições, buscando padrões que acreditam ter detectado. Se suas previsões estiverem corretas, então elas podem mudar para a sequência recém-descoberta.

Não culparia um participante por continuar formulando hipóteses — afinal, como poderiam saber que a sequência está além de sua capacidade de previsão? Mas culparia um participante por apostar nas suposições, quando isso não é necessário para coletar informações, especialmente quando centenas de suposições anteriores já foram refutadas.

Até mesmo um ser humano pode ser tão confiante?

Suspeitaria que algo mais simples estivesse acontecendo — que a estratégia totalmente azul simplesmente não ocorreu aos participantes.

As pessoas veem uma mistura de cartas, principalmente azuis, com algumas vermelhas, e supõem que a estratégia de aposta ideal deva ser uma mistura de cartas principalmente azuis com algumas vermelhas.

É uma ideia contraintuitiva que, dada a informação incompleta, a estratégia de aposta ideal não se assemelhe a uma sequência típica de cartas.

É uma ideia contraintuitiva que a estratégia ideal é seguir a lei, mesmo em um ambiente com elementos aleatórios.

Parece que seu comportamento deveria ser imprevisível, assim como o ambiente — mas não! Uma chave aleatória não abre um cadeado aleatório só porque são “ambos aleatórios”.

Você não combate fogo com fogo; você combate o fogo com água. Mas esse pensamento envolve uma etapa extra, um novo conceito não diretamente ativado pela declaração do problema e, portanto, não é a primeira ideia que vem à mente.

No dilema das cartas azuis e vermelhas, nossa compreensão parcial nos diz — a cada rodada — que a melhor escolha é a carta azul. Este conselho de nossa compreensão parcial é o mesmo em cada rodada. Se desafiarmos nossa compreensão parcial por 30% do tempo e escolhermos a carta vermelha, faremos pior — porque agora estamos sendo completamente estúpidos, apostando no que sabemos ser o resultado menos provável.

Se você escolher a carta vermelha a cada rodada, você se sairá tão mal quanto poderia; você seria 100% estúpido. E, se escolher a carta vermelha, 30% das vezes, enfrentando 30% de cartas vermelhas, estará sendo 30% estúpido.

Quando sua compreensão é incompleta — o que significa que o mundo parece ter um elemento de aleatoriedade — randomizar suas ações, não resolve o problema. A randomização de suas ações leva você para mais longe do alvo, não para mais perto. Em um mundo já incerto, abandonar sua inteligência só piora as coisas.

É uma ideia contraintuitiva que a estratégia ideal possa ser pensar de forma lógica, mesmo sob condições de incerteza.

E assim, poucos são racionalistas, pois a maioria que percebe um mundo caótico tentará combater o caos com o caos. Você tem que dar um passo extra e pensar em algo que não vem à sua mente, para imaginar combater o fogo com algo que não é fogo em si.

Você já ouviu os não iluminados dizerem: “A racionalidade funciona bem para lidar com pessoas racionais, mas o mundo não é racional”. No entanto, diante de um oponente irracional, abandonar sua própria razão não o ajudará. Existem maneiras lógicas de pensar que ainda produzem a melhor resposta, mesmo quando confrontadas com um oponente que viola essas leis. A teoria da decisão não se desintegra em chamas e morre quando confrontada com um oponente que a desobedece.

Isso não é mais óbvio do que a ideia de apostar tudo azul, diante de uma sequência de cartas azuis e vermelhas. Mas cada aposta que você faz no vermelho é uma perda esperada, e o mesmo ocorre com cada afastamento do Caminho em seu próprio pensamento.

Quantos episódios de Jornada nas Estrelas são refutados assim? E, quantas teorias de IA?

Referências

1. *Dawes, Rational Choice in An Uncertain World; Yaacov Schul and Ruth Mayo, "Searching for Certainty in an Uncertain World: The Difficulty of Giving Up the Experiential for the Rational Mode of Thinking," Journal of Behavioral Decision Making 16, no. 2 (2003): 93–106, doi:10.1002/bdm.434.*
2. *Amos Tversky and Ward Edwards, "Information versus Reward in Binary Choices," Journal of Experimental Psychology 71, no. 5 (1966): 680–683, doi:10.1037/h0023123.*

40 — Minha juventude selvagem e imprudente

Dizem que os pais fazem tudo o que dizem aos filhos para não fazer, e é dessa forma que eles aprendem o que não devem fazer.

Há muito tempo, em um passado distante, eu era um racionalista tradicional dedicado e me considerava habilidoso nesse campo, mas ainda não conhecia o Caminho de Bayes. Quando jovem, Eliezer se deparou com uma questão aparentemente misteriosa e, apesar de seguir os preceitos da Racionalidade Tradicional, não conseguiu evitar inventar uma Resposta Misteriosa. Até hoje, esse é o erro mais embaraçoso que cometi em minha vida e ainda sinto arrepios ao pensar nele.

Qual foi a minha resposta misteriosa para uma pergunta misteriosa? Eu não vou descrevê-la, pois seria uma história longa e complicada. Eu era jovem e apenas um racionalista tradicional que não conhecia os ensinamentos de Tversky e Kahneman. Eu sabia sobre a Navalha de Ocam, mas não sobre a falácia da conjunção. Pensei que poderia me safar pensando coisas complicadas , no estilo literário dos pensamentos complicados que leio nos livros de ciências, sem perceber que a complexidade correta só é possível quando cada etapa é definida de forma clara e precisa. Hoje, um dos principais conselhos que dou aos jovens aspirantes a racionalistas é: “Não tente longas cadeias de raciocínio ou planos complicados”.

Nada mais precisa ser dito: mesmo após inventar minha “resposta”, o fenômeno ainda era um mistério para mim e possuía a mesma qualidade assombrosa de impenetrabilidade que tinha no início.

Não se engane, o jovem Eliezer não era estúpido. Todos os erros que ele cometeu ainda são cometidos hoje por cientistas respeitados em revistas de renome. Seria necessário um discernimento mais refinado para protegê-lo do que o que foi ensinado como racionalista tradicional.

De fato, o jovem Eliezer seguiu diligente e meticulosamente as injunções da Racionalidade Tradicional ao se desviar do caminho.

Como um racionalista tradicional, o jovem Eliezer teve o cuidado de garantir que sua Resposta Misteriosa fizesse uma previsão ousada da experiência futura. Ele esperava que os futuros neurologistas descobrissem que os neurônios estavam explorando a gravidade quântica, semelhante à ideia de Sir Roger Penrose. Isso exigia que os neurônios mantivessem um certo grau de coerência quântica, algo que poderia ser encontrado ou não.

Mas a minha hipótese não fez previsões retrospectivas. Segundo a Ciência Tradicional, previsões retrospectivas são inválidas, então por que se preocupar em fazê-las? Por outro lado, para um bayesiano, se uma hipótese não tem uma razão de verossimilhança favorável em relação à “não sei”, isso suscita a questão de por que acreditar hoje em algo mais complicado do que “não sei”. No entanto, naquela época, eu não conhecia o Caminho de Bayes, então não estava pensando em razões de verossimilhança ou me concentrando na densidade de probabilidade. Eu havia feito uma previsão refutável; essa não era a lei?

Como um Racionalista Tradicional, o jovem Eliezer teve o cuidado de não acreditar em magia, misticismo, chauvinismo do carbono ou qualquer coisa do tipo. Eu professava com orgulho a minha Resposta Misteriosa: “É apenas física, como o resto da física!” Como se pudesse salvar a magia de ser um isomorfo cognitivo da magia, chamando-a de gravidade quântica. No entanto, naquela época, eu não conhecia o Caminho de Bayes e não via o nível em que minha ideia era isomórfica à magia. Eu havia dado minha lealdade à física, mas isso não me salvou; o que a teoria da probabilidade sabe sobre alianças? Eu evitava tudo o que a Racionalidade Tradicional me dizia para evitar, mas o que sobrava ainda era mágico.

Sem dúvida, a minha lealdade à Racionalidade Tradicional me ajudou a sair do buraco em que me enfiei. Se eu não fosse um Racionalista Tradicional, estaria completamente ferrado. No entanto, a Racionalidade Tradicional não foi suficiente para me corrigir. Isso me levou a erros diferentes daqueles que eu havia explicitamente proibido.

Quando penso em como o meu eu mais jovem seguiu cuidadosamente as regras da Racionalidade Tradicional e, ainda assim, errou, isso explica por que as pessoas que se autodenominam "racionalistas" não governam o mundo. Você precisa de muita racionalidade antes que ela possa fazer qualquer coisa além de levá-lo a novos e interessantes erros.

A Racionalidade Tradicional é ensinada como uma arte, e não como uma ciência; você lê a biografia de físicos famosos, descrevendo as lições que a vida lhes ensinou, e tenta seguir seus conselhos. Mas você não viveu a vida deles, e metade do que eles tentam descrever é um instinto forjado em suas experiências.

Do jeito que a Racionalidade Tradicional é concebida, seria aceitável que eu desperdiçasse trinta anos com minha ideia tola, contanto que eu eventualmente conseguisse refutá-la e fosse honesto comigo mesmo sobre o que minha teoria previa, aceitando a refutação quando chegasse, e assim por diante. Isso é suficiente para fazer a Roda da Ciência girar, mas é um pouco difícil para aqueles que perdem trinta anos de suas vidas. A Racionalidade Tradicional é uma jornada, não uma dança. Ela foi projetada para te levar eventualmente à verdade, oferecendo a você muito tempo para apreciar o caminho.

Os racionalistas tradicionais podem concordar em discordar. A Racionalidade Tradicional não considera o pensamento como uma arte exata, na qual existe apenas uma estimativa de probabilidade correta, dada a evidência. Na Racionalidade Tradicional, formular hipóteses e testá-las. Mas a experiência me ensinou que, se você não sabe e adivinha, acabará se enganando.

O Caminho de Bayes também é uma arte imprecisa, pelo menos do jeito que estou falando sobre ele. Esses ensaios ainda são tentativas desajeitadas de traduzir em palavras as lições que seriam mais bem-ensinadas pela experiência. Mas pelo menos há matemática subjacente, além de evidências experimentais da psicologia cognitiva sobre como os humanos realmente pensam. Talvez isso seja o suficiente para cruzar o limiar estratosférico necessário para uma disciplina que te permite realmente acertar, em vez de apenas te levar a cometer novos erros interessantes.

41 — Falhando em aprender com a História

Havia uma época, durante minha juventude selvagem e imprudente, quando eu ainda não conhecia o Caminho de Bayes, e respondi a uma pergunta aparentemente misteriosa com uma Resposta Misteriosa. Muitas falhas ocorreram em sequência, mas um erro se destaca como o mais crítico: meu eu mais jovem não percebeu que resolver um mistério deveria torná-lo menos confuso. Eu estava tentando explicar um Fenômeno Misterioso, que para mim significava fornecer uma causa para ele, encaixando-o em um modelo integrado de realidade. Mas por que isso deveria tornar o fenômeno menos misterioso, quando essa é sua natureza? Eu estava tentando explicar o Fenômeno Misterioso, não o transformar (por alguma alquimia impossível) em um fenômeno mundano, um fenômeno que nem mesmo exigiria uma explicação incomum em primeiro lugar.

Como Racionalista Tradicional, eu conhecia as histórias dos astrólogos e da astronomia, dos alquimistas e da química, dos vitalistas e da biologia. Mas o Fenômeno Misterioso não era assim. Era algo novo, algo estranho, algo mais difícil, algo que a ciência comum não conseguiu explicar por séculos…

— Como se as estrelas, a matéria e a vida não tivessem sido mistérios por centenas e milhares de anos, desde o surgimento do pensamento humano até a ciência finalmente resolvê-los —

Aprendemos sobre astronomia, química e biologia na escola, e parece-nos que esses assuntos sempre foram o domínio apropriado da ciência, que nunca foram misteriosos. Quando a ciência ousa desafiar um novo Grande Enigma, as crianças daquela geração ficam céticas, pois nunca viram a ciência explicar algo que lhes parecesse misterioso. A ciência só possibilita explicar assuntos científicos, como estrelas, matéria e vida.

Eu achava que a lição da história era que astrólogos, alquimistas e vitalistas tinham uma falha de caráter inata, uma tendência ao misteriosismo, que os levava a apresentar explicações misteriosas para assuntos não misteriosos. Mas certamente, se um fenômeno realmente era muito estranho, uma explicação estranha poderia estar em ordem?

Foi somente depois, quando comecei a ver a estrutura mundana no mistério, que percebi em que lugar estava. Só então percebi o quão razoável o vitalismo parecia na época, o quão surpreendente e embaraçosa havia sido a resposta do universo: “A vida é mundana e não precisa de uma explicação estranha”.

Lemos a história, mas não a experimentamos, não a vivenciamos. Se eu tivesse pessoalmente conjecturado os mistérios da astrologia e, em seguida, descoberto a mecânica newtoniana; conjecturado os mistérios da alquimia e, em seguida, descoberto a química; conjecturado os mistérios do vitalismo e, em seguida, descoberto a biologia, eu teria me lembrado da Resposta Misteriosa e dito para mim mesmo: “Eu não cairei nessa armadilha novamente.”

42 — Disponibilizando a História

Existe um hábito de pensamento que chamo de falácia lógica da generalização a partir de evidências fictícias. Por exemplo, jornalistas que falam sobre os filmes do Exterminador do Futuro em uma reportagem sobre IA não costumam tratá-los como profecia ou verdade. No entanto, o filme é lembrado — e disponibilizado, por assim dizer — como se fosse um caso histórico ilustrativo. É como se os jornalistas tivessem visto isso acontecer em algum outro planeta, e que, portanto, pode muito bem acontecer aqui. Mais informações sobre esse tema podem ser encontradas na Seção 7 do artigo Cognitive biases potentially affecting judgment of global risks (Vieses Cognitivos que Podem Afetar o Julgamento de Riscos Globais). [1]

Há também um erro inverso que consiste em não se mover suficientemente por evidências históricas. O problema em generalizar a partir de evidências fictícias é que estas nunca aconteceram de verdade. Elas não são extraídas da mesma distribuição do nosso universo real, e a ficção difere da realidade de maneiras sistemáticas. Por outro lado, a história aconteceu e deve estar disponível para consulta. Em nosso ambiente ancestral, não havia cinema e, portanto, tudo o que você via com seus próprios olhos era verdade. É de se admirar que as ficções que vemos em imagens realistas em movimento, tenham um impacto tão grande sobre nós? Em contrapartida, as coisas que realmente aconteceram foram registradas apenas em tinta no papel, e nunca as testemunhamos acontecer.

Não nos lembramos de tê-las vivenciado.

O erro inverso consiste em tratar a história como algo irrelevante, processando-a com a mesma parte da mente que lida com os romances que você lê. Pode-se afirmar que a história é verdadeira em vez de ficção, mas isso não significa que você esteja sendo tão movido quanto deveria. Muitos vieses cognitivos envolvem uma insuficiência na reação às informações secas e abstratas.

Era uma vez… dei uma resposta misteriosa a uma pergunta misteriosa, sem perceber que estava cometendo o mesmo erro que os astrólogos que concebem explicações místicas para as estrelas, ou alquimistas que concebem propriedades mágicas da matéria, ou vitalistas que postulam um "elã vital" opaco para explicar tudo de biologia.

Quando finalmente percebi onde estava, houve um súbito choque de conexão inesperada com o passado. Percebi que a invenção e destruição do vitalismo — sobre o qual só havia lido em livros — aconteceu realmente com pessoas reais, que vivenciaram isso da mesma forma que experimentei a invenção e destruição da minha própria resposta misteriosa. E também percebi que, se eu tivesse realmente experimentado o passado — se eu mesmo tivesse vivido as revoluções científicas do passado, em vez de lê-las nos livros de história — provavelmente não teria cometido o mesmo erro novamente. Eu não teria dado outra resposta misteriosa; as primeiras mil lições teriam martelado a moral.

Então, pensei que, para sentir suficientemente a força da história, eu deveria tentar aproximar os pensamentos de um Eliezer que viveu através da história — eu deveria tentar pensar como se tudo o que li nos livros de história tivesse realmente acontecido comigo. (Com reponderação apropriada para o viés de disponibilidade dos livros de história — devo me lembrar de ser mil camponeses para cada governante.) Eu deveria mergulhar na história, imaginar viver eras que eu só via como tinta no papel.

Por que eu deveria me lembrar do primeiro voo dos irmãos Wright? Eu não estava lá. Mas como racionalista, eu poderia ousar não lembrar quando o evento realmente aconteceu? Existe tanta diferença entre ver um evento através dos seus olhos — que, na verdade, é uma cadeia causal envolvendo fótons refletidos, não uma conexão direta — e ver um evento por meio de um livro de história? Fótons e livros de história descendem por cadeias causais do próprio evento.

Tive que superar a falsa amnésia de ter nascido em um determinado momento. Eu tinha que relembrar — disponibilizar — todas as memórias, não apenas as memórias que, por coincidência, pertenciam a mim e à minha época.

A Terra envelheceu, de repente.

Na minha memória anterior, sempre existiram os Estados Unidos — nunca houve um tempo sem existir. Eu não havia me lembrado, até aquele momento, de como o Império Romano surgiu, trouxe paz e ordem, e durou tantos séculos. Eu havia esquecido que as coisas nem sempre foram assim; no entanto, o Império caiu, os bárbaros invadiram minha cidade e todo o conhecimento que eu possuía foi perdido. O mundo moderno parecia mais frágil aos meus olhos, mas ele não foi o primeiro mundo moderno.

Cometi tantos erros repetidamente, simplesmente porque não me lembrava de tê-los cometido em tantas outras ocasiões.

É curioso como às vezes as pessoas se perguntam se superar preconceitos é importante. Você não se lembra quantas vezes seus preconceitos o prejudicaram? Percebi que a amnésia repentina muitas vezes segue um erro fatal. Mas acredite em mim, isso aconteceu. Eu me lembro; eu não estava lá.

Então, da próxima vez que você duvidar do futuro e de suas estranhezas, lembre-se de que você nasceu em uma tribo de caçadores-coletores há dez mil anos, quando ninguém sabia nada sobre ciência. Lembre-se do choque, quando a ciência explicou os grandes e terríveis mistérios sagrados que você costumava reverenciar tanto. Lembre-se de como você acreditava que poderia voar ao comer cogumelos específicos, mas depois teve que aceitar com decepção que isso nunca aconteceria, até que um dia você finalmente voou. Lembre-se da convicção de que a escravidão era justa e adequada, até mudar de ideia. Não se esqueça de que você não imaginou essa mudança. A verdade é que você não imaginou. Lembre-se de como, século após século, o mundo mudou de maneiras que você nunca imaginou.

Assim, talvez você fique menos chocado com o que acontecerá a seguir.

Referências

1. *Eliezer Yudkowsky, "Cognitive Biases Potentially Affecting Judgment of Global Risks," in Global Catastrophic Risks, ed. Nick Bostrom and Milan M. Ćirković (New York: Oxford University Press, 2008), 91–119.*

43 — Explicar, adorar, ignorar?

Enquanto nossa tribo caminha pelas pastagens em busca de frutas e presas, às vezes acontece de a água cair do céu.

"Pergunto ao sábio barbudo de nossa tribo: "Por que a água cai do céu às vezes?"

Ele pensa por um momento, essa pergunta nunca lhe ocorreu antes, e então diz: "Ocasionalmente, os espíritos do céu lutam e, quando o fazem, seu sangue pinga do céu."

"Pergunto: De onde vêm os espíritos do céu?"

Sua voz diminui para um sussurro. "Desde antes do tempo. Desde muito, muito tempo atrás."

Quando chove e você não sabe o porquê, você tem várias opções. Primeiro, você poderia simplesmente não perguntar o porquê — não dar continuidade à pergunta ou nunca pensar nela em primeiro lugar. Este é o comando "Ignorar", que o sábio barbudo originalmente escolheu. Em segundo lugar, você pode tentar inventar alguma explicação, o comando "Explicar", como o homem barbudo fez em resposta à sua primeira pergunta. Em terceiro lugar, você pode desfrutar da sensação de mistério — o comando de Adoração.

Agora, como você deve perceber nesta história, cada vez que você seleciona Explicar, o melhor cenário é obter uma explicação, como "espíritos do céu". Mas essa explicação em si está sujeita ao mesmo dilema — Explicar, Adorar ou Ignorar? Cada vez que você clica em Explicar, a ciência trabalha por um tempo, retorna uma explicação e, em seguida, outra caixa de diálogo aparece. Como bons racionalistas, nos sentimos obrigados a continuar clicando em Explicar, mas parece uma estrada sem fim.

Você clica em Explicar pelo resto de sua vida e obtém química; você clica em Explicar para química e obtém átomos; você clica em Explicar para átomos e obtém elétrons e núcleos; você clica em Explicar para núcleos e obtém cromodinâmica quântica e quarks; você clica em Explicar para quarks e obtém o Big Bang…

Podemos clicar em "Explicar" para o Big Bang e esperar enquanto a ciência avança em seu processo, e talvez um dia ela retorne com uma explicação perfeitamente boa. Mas isso apenas abrirá outra caixa de diálogo. Então, se continuarmos o suficiente, devemos chegar a uma caixa de diálogo especial, uma nova opção, uma explicação que não precisa de explicação, um lugar onde a cadeia termina — e essa, talvez, seja a única explicação que vale a pena conhecer.

Pronto, acabei de clicar em "Adoração".

Nunca se esqueça de que há muito mais maneiras de adorar algo do que acender velas ao redor de um altar.

Se eu tivesse dito: "Hum, isso parece paradoxal. Eu me pergunto: como o aparente paradoxo é resolvido?". Então eu teria clicado em "Explicar", o que às vezes demora um pouco para produzir uma resposta.

E se todo o assunto lhe parece sem importância, ou irrelevante, ou se você prefere deixar para pensar nisso até amanhã, então clique em "Ignorar".

Selecione sua opção com sabedoria.

44 — “Ciência” como freio da curiosidade

Imagine que eu, diante das câmeras de televisão ao vivo, levantei minhas mãos e cantei abracadabra, fazendo uma luz brilhante surgir no espaço vazio entre as minhas mãos estendidas. Agora, imagine que executei esse ato de feitiçaria flagrante e inconfundível sob a total supervisão de James Randi e de todos os céticos presentes. A maioria das pessoas, eu imagino, ficaria bastante curiosa para saber o que estava acontecendo.

Mas, agora, suponhamos que eu não apareça na televisão. Eu não desejo compartilhar o poder, nem a verdade por trás dele. Quero manter minha feitiçaria em segredo, mas também quero conjurar meus feitiços quando e onde eu quiser. Quero lançar meu brilhante clarão de luz para poder ler um livro no trem sem que ninguém fique curioso. Existe um feitiço que acaba com a curiosidade?

Sim, de fato existe! Sempre que alguém me pergunta: “Como você fez isso?”, eu apenas respondo: “Ciência!”

Essa resposta não é uma explicação real, mas sim um freio da curiosidade. Ela não diz se a luz irá aumentar ou diminuir, mudar de cor ou saturação e certamente não diz como fazer você mesmo uma luz semelhante. Na verdade, você não sabe nada mais do que sabia antes de eu dizer a palavra mágica. Mas, ainda assim, você fica satisfeito por não estar acontecendo nada incomum.

Melhor ainda, o mesmo truque funciona com um interruptor de luz padrão. Basta apertar um interruptor e uma lâmpada acende. Por quê?

Na escola, aprendemos que a senha da lâmpada é “Eletricidade!” A esta altura, espero que você esteja cauteloso em marcar a lâmpada como “entendida” com base nisso. Será que dizer “Eletricidade!” permite que você faça cálculos que controlarão sua antecipação da experiência? Há, pelo menos, muito mais a aprender. (Os físicos devem ignorar este parágrafo e substituí-lo por um problema na teoria evolutiva, onde a substância da teoria está novamente em cálculos que poucas pessoas sabem fazer.)

Se você pensasse que a lâmpada era cientificamente inexplicável, ela capturaria toda a sua atenção. Você abandonaria qualquer outra coisa que estivesse fazendo e se concentraria naquela lâmpada.

Mas o que significa a expressão “cientificamente explicável”? Significa que outra pessoa sabe como a lâmpada funciona. Quando você é informado de que a lâmpada é “cientificamente explicável”, você não sabe mais do que antes; você não sabe se a lâmpada vai aumentar ou diminuir a luz. Mas o fato de outra pessoa saber desvaloriza o seu conhecimento. Sua curiosidade diminui. Alguém poderia argumentar: “Se a lâmpada fosse desconhecida da ciência, você poderia ganhar fama e fortuna investigando-a”. No entanto, não se trata de ganância ou ambição profissional. Estou falando da emoção crua, da curiosidade, da sensação de intriga. Por que sua curiosidade deveria ser diminuída porque outra pessoa sabe como a lâmpada funciona? Isso não é inveja? Não é suficiente que você saiba? Outras pessoas também devem ser ignorantes para que você seja feliz?

Existem bens para os quais o conhecimento pode servir além da curiosidade, como a utilidade social da tecnologia. Para esses bens instrumentais, é importante saber se outra pessoa já tem o conhecimento no espaço local. Mas, para minha própria curiosidade, por que isso importa?

Além disso, considere as consequências de permitir que “Alguém já sabe a resposta” funcione como um obstáculo à curiosidade. Imagine que um dia você entra em sua sala de estar e vê um elefante verde gigante, aparentemente flutuando no ar, cercado por uma aura de luz prateada.

“O que diabos é isso?” você pergunta.

E uma voz vem do alto do elefante, dizendo:

“ALGUÉM JÁ SABE POR QUE ESTE ELEFANTE ESTÁ AQUI”.

“Ah”, você responde, “nesse caso, não importa”, e segue para a cozinha.

Não conheço a grande teoria unificada das leis da física deste universo. Também não sei muito sobre anatomia humana, exceto pelo cérebro. Não consigo apontar em meu corpo onde estão meus rins e não consigo me lembrar imediatamente o que meu fígado faz. (Não me orgulho disso. Infelizmente, com toda a matemática que preciso estudar, provavelmente não aprenderei anatomia tão cedo.)

Em relação à curiosidade, devo me sentir mais intrigado com minha ignorância das leis fundamentais da física do que com o fato de não saber muito sobre o que acontece dentro do meu próprio corpo?

Se eu levantasse as minhas mãos e lançasse um feitiço de luz, você ficaria intrigado. Mas, será que você ficaria menos intrigado com o simples fato de eu ter levantado as mãos? Quando você levanta o braço e acena com a mão, esse ato de vontade é coordenado (entre outras áreas do cérebro) pelo cerebelo. Aposto que você não sabe como o cerebelo funciona. Sei um pouco — embora apenas os detalhes grosseiros, não o suficiente para realizar cálculos. E daí? O que isso importa, se você não sabe? Por que haveria um padrão duplo de curiosidade para feitiçaria e movimentos de mão?

Olhe-se no espelho. Você sabe o que está olhando? Sabe o que parece por trás dos seus olhos? Você sabe quem é? A ciência tem respostas para algumas dessas perguntas, mas para outras, não. Mas por que essa distinção deveria interessar à sua curiosidade?

Você sabe como funcionam os joelhos? Você sabe como os seus sapatos foram feitos? Você sabe por que o monitor do seu computador brilha? Você sabe por que a água é molhada?

O mundo ao seu redor está cheio de enigmas. Priorize, se necessário. Mas não se queixe que a ciência cruel esvaziou o mundo de mistérios. Com esse tipo de raciocínio, eu poderia fazer você ignorar um elefante em sua sala de estar.

45 — Verdadeiramente parte de você

Um artigo clássico de Drew McDermott, intitulado *Artificial Intelligence Meets Natural Stupidity* (A inteligência artificial se encontra com a estupidez natural), criticou programas de inteligência artificial que tentam representar conceitos como "felicidade é um estado de espírito" usando uma rede semântica. [1]

ESTADO DE ESPÍRITO

^

| É-UM

| FELICIDADE

E, é claro, não há nada no nodo **FELICIDADE**; é apenas um token LISP nu com um nome sugestivo em inglês.

Então, McDermott diz: "Um bom teste para o programador disciplinado é tentar usar gensyms em lugares-chave e ver se ele ainda admira seu sistema. Por exemplo, se **ESTADO-DE-ESPÍRITO** for renomeado para G1073…" então teríamos É-UMA (FELICIDADE, G1073), "o que parece muito mais duvidoso".

Ou seja, se você substituísse símbolos aleatórios por todos os nomes sugestivos em inglês, seria completamente incapaz de descobrir o que G1071(G1072, G1073) significava. O programa de IA foi feito para representar hambúrgueres? Maçãs? Felicidade? Quem sabe? Se você excluir os nomes sugestivos em inglês, eles não voltarão a aparecer.

Suponha que um físico lhe diga que "a luz são ondas" e você acredite nele. Agora você tem uma pequena rede em sua cabeça que diz:

É-UMA (LUZ, ONDAS).

Se alguém lhe perguntar "Do que é feita a luz?", você pode responder "Ondas!"

Como McDermott diz: "O desafio é fazer com que o ouvinte perceba o que foi dito. Não 'entenda', mas 'observe'." Agora, imagine que, em vez disso, um físico dissesse: "A luz é feita de pequenas coisas curvas" (o que, a propósito, não é verdade). Você notaria alguma diferença na experiência antecipada?

Como você pode perceber que não deve confiar em seu aparente conhecimento de que "a luz são ondas"? Um teste que você pode aplicar é perguntar: "Eu poderia regenerar esse conhecimento se, de alguma forma, ele fosse excluído da minha mente?"

Isso é semelhante em espírito a embaralhar os nomes dos tokens LISP com nomes sugestivos em seu programa de IA e ver se outra pessoa consegue descobrir a que eles supostamente "se referem". Também é semelhante em espírito a observar que um Aritmético Artificial programado para gravar e reproduzir

não pode regenerar o conhecimento se você o excluir da memória, até que outro humano o insira novamente no banco de dados. Assim como se você esquecesse que "a luz são ondas", você não poderia recuperar o conhecimento, exceto da mesma maneira que obteve o conhecimento para começar — perguntando a um físico. Você não poderia gerar o conhecimento para si mesmo, da maneira que os físicos o geraram originalmente.

As mesmas experiências que nos levam a formular uma crença, conectam essa crença a outros conhecimentos e entradas sensoriais e motoras. Se você vir um castor mastigando uma tora, então você saberá como é essa coisa-que-mastiga-a-tora e poderá reconhecê-la em ocasiões futuras, independentemente de ser chamada de "castor" ou não. Mas se você adquirir suas crenças sobre castores por alguém contando fatos sobre "castores", você pode não conseguir reconhecer um castor quando vir um.

Este é o terrível perigo de tentar ensinar fatos a uma Inteligência Artificial que ela não pode aprender por si mesma. É também o terrível perigo de tentar ensinar alguém sobre física que eles não podem verificar por si mesmos. Pois o que os físicos querem dizer com "onda" não é "coisinha ondulada", mas um conceito puramente matemático.

Como observado por Davidson, se você acredita que "castores" vivem em desertos, são completamente brancos e pesam 136 quilos quando adultos, então você não tem nenhuma crença sobre castores, seja ela verdadeira ou falsa. Sua crença sobre "castores" não é precisa o suficiente para estar incorreta. Se você não tem experiência suficiente para regenerar crenças quando elas são apagadas, então você tem experiência suficiente para conectá-las a algo? Wittgenstein disse: "Uma roda que pode girar sem mover mais nada com ela não faz parte do mecanismo".

Quando comecei a ler sobre Inteligência Artificial — antes mesmo de ler McDermott — percebi que seria uma boa ideia sempre me perguntar: "Como eu poderia regenerar esse conhecimento se ele fosse apagado da minha mente?"

Quanto mais profunda for a exclusão, mais rigoroso será o teste. Se todas as provas do Teorema de Pitágoras fossem apagadas da minha mente, eu conseguiria prová-lo? Acredito que não. Se todo o conhecimento do Teorema de Pitágoras fosse deletado da minha mente, eu notaria o Teorema de Pitágoras para prová-lo? Isso é mais difícil de afirmar sem o colocar à prova; mas se me fosse dado um triângulo retângulo com lados de comprimento 3 e 4 e me dissessem que o comprimento da hipotenusa é calculável, acho que poderia calculá-lo, se ainda soubesse todo o resto da minha matemática.

E quanto à noção de prova matemática? Se ninguém nunca tivesse me falado sobre isso, eu conseguiria inventá-la com base em outras crenças que tenho? Houve um tempo em que a humanidade não possuía tal conceito. Alguém teve que inventá-lo. O que eles perceberam? Eu perceberia se visse algo igualmente novo e igualmente importante? Poderia pensar tão fora da caixa?

Quanto do seu conhecimento você conseguiria regenerar? Qual é a profundidade da sua exclusão? Não é apenas um teste para descartar crenças insuficientemente conectadas. É uma forma de absorver uma fonte de conhecimento, não apenas um fato.

Um pastor constrói um sistema de contagem que funciona jogando uma pedrinha em um balde sempre que uma ovelha sai do redil e retirando uma pedrinha sempre que uma ovelha retorna. Se você, como aprendiz, não entender esse sistema — se for uma mágica que funciona sem motivo aparente — então, você não saberá o que fazer se acidentalmente deixar cair uma pedrinha a mais no balde. Aquilo que você não pode fazer sozinho, você não pode refazer quando a situação exigir. Você não pode voltar para a fonte, ajustar uma das configurações de parâmetro e regenerar a saída sem a fonte. Se "dois mais quatro é igual a seis" é um fato para você, e então um dos elementos muda para "cinco", como você saberá que "dois mais cinco é igual a sete" quando simplesmente lhe foi dito que "dois mais quatro é igual a seis"?

Se você vir uma pequena planta que solta uma semente toda vez que um pássaro passa por ela, não ocorrerá a você que pode usar essa planta para automatizar parcialmente o contador de ovelhas. Embora você tenha aprendido algo que o criador original poderia usar para melhorar sua invenção, você não pode voltar à fonte e recriá-la.

Quando você é a fonte de um pensamento, esse pensamento pode evoluir com você, à medida que você adquire novos conhecimentos e habilidades. Quando você é a fonte de um pensamento, ele se torna parte verdadeiramente de você e cresce com você. Esforce-se para ser a fonte de todos os pensamentos que valem a pena ser pensados. Se o pensamento veio originalmente de fora, certifique-se de que também vem de dentro. Pergunte-se continuamente: "Como eu regeneraria esse pensamento se ele fosse apagado?" Quando tiver uma resposta, imagine que esse conhecimento também foi apagado. E quando você encontrar uma fonte, veja o que mais ela pode oferecer.

Referências

1. *Drew McDermott, "Artificial Intelligence Meets Natural Stupidity," SIGART Newsletter, no. 57 (1976): 4–9, doi:10.1145/1045339.1045340.*
2. *Richard Rorty, "Out of the Matrix: How the Late Philosopher Donald Davidson Showed That Reality Can't Be an Illusion," The Boston Globe (October 2003).*

Interlúdio: verdade simples

Lembro-me de um artigo que escrevi sobre o existencialismo. Minha professora me devolveu com um F. Ela sublinhou a palavra "verdade" todas as vezes que apareceu no texto, cerca de vinte vezes, acompanhadas de um ponto de interrogação ao lado de cada uma delas. Ela queria entender o que eu realmente queria dizer com "verdade"[16].

— Danielle Egan, jornalista

Este ensaio visa restaurar uma visão simples da verdade.

Imagine alguém lhe dizendo: "Meu óleo de cobra milagroso pode curar seu câncer de pulmão em apenas três semanas". Você responde: "Mas um estudo clínico não comprovou que essa afirmação é falsa?". A pessoa retorna: "Essa ideia de 'verdade' é bastante ingênua; o que você quer dizer com 'verdadeiro'?"

Muitas pessoas, quando questionadas, não conseguem explicar com detalhes rigorosos o que é a verdade. No entanto, elas não deveriam abandonar o conceito de "verdade". Houve um tempo em que ninguém conhecia as equações da gravidade em detalhes rigorosos, mas se você caísse de um penhasco, você cairia.

Frequentemente, vejo — especialmente em listas de discussão na internet — que, em meio a outras conversas, alguém diz "X é verdadeiro" e, em seguida, surge uma discussão sobre o uso da palavra "verdadeiro". Este ensaio não almeja ser uma referência enciclopédica para esse argumento. Em vez disso, espero que os debatedores leiam este ensaio e depois voltem ao que estavam discutindo antes que alguém questione a natureza da verdade.

Neste ensaio, faço perguntas. Se você vir o que parece uma resposta óbvia, é provavelmente a resposta que procuro. A escolha óbvia nem sempre é a melhor escolha, mas às vezes, caramba, é. Não paro de procurar assim que encontro uma resposta óbvia, mas se continuo procurando e a resposta aparentemente óbvia ainda parece óbvia, não me sinto culpado por mantê-la. Ah, claro, todo mundo pensa que dois mais dois são quatro, todo mundo diz que dois mais dois são quatro e, na mera labuta mundana da vida cotidiana, todos se comportam como se dois mais dois fossem quatro, mas o que realmente são dois mais dois? Tanto quanto posso imaginar, quatro. Ainda são quatro, mesmo que eu repita a pergunta em um tom de voz solene e imponente. Muito simples, você diz? Talvez, nesta ocasião, a vida não precise ser complicada. Isso não seria revigorante?

Se você é uma daquelas pessoas sortudas para quem a pergunta parece trivial desde o início, espero que ainda pareça trivial no final. Se você se sentir confuso com questões profundas e significativas, lembre-se de que, se você sabe exatamente como um sistema funciona e pode construir um você mesmo com baldes e pedrinhas, isso não deve ser um mistério para você.

16 NT: Tradução livre do texto original em inglês. *I remember this paper I wrote on existentialism. My teacher gave it back with an F. She'd underlined true and truth wherever it appeared in the essay, probably about twenty times, with a question mark beside each. She wanted to know what I meant by truth.*

Se a interpretação de uma metáfora como tal ameaçar confundir você, tente interpretar tudo literalmente.

Imagine que, em uma época anterior à história registrada ou à matemática formal, eu sou um pastor que tem dificuldades em rastrear suas ovelhas. Elas dormem em um curral cercado, alto o suficiente para protegê-las dos lobos que vagam à noite. Diariamente, solto as ovelhas para pastar e, todas as noites, eu as encontro e as devolvo ao curral. Se alguma ovelha ficar do lado de fora, seu corpo será encontrado na manhã seguinte, morto e meio comido por lobos. Mas é desanimador passar horas procurando por uma ovelha perdida quando sei que provavelmente todas estão no curral. Às vezes, eu desisto cedo e geralmente estou certo, mas cerca de uma em cada dez vezes, uma ovelha está morta na manhã seguinte.

Se apenas houvesse uma maneira de saber se todas as ovelhas continuam pastando, sem a inconveniência de olhar! Eu tentei vários métodos: joguei os bastões de adivinhação da minha tribo; treinei meus poderes psíquicos para localizar as ovelhas através da clarividência; e procurei cuidadosamente razões para acreditar que todas as ovelhas estavam no curral. Não importa o que eu faça, cerca de uma em cada dez vezes que dormirei, achando que todas as ovelhas estão seguras, encontrarei uma ovelha morta na manhã seguinte. Talvez eu perceba que meus métodos não estão funcionando e me desculpe cuidadosamente por cada falha, mas meu dilema permanece o mesmo. Posso passar uma hora procurando em todos os lugares possíveis, quando na maioria das vezes não há ovelhas restantes, ou posso dormir cedo e perder, em média, uma ovelha em cada dez.

No final da tarde, você está especialmente cansado. Você joga os bastões de adivinhação e eles indicam que todas as ovelhas voltaram. Você visualiza cada canto e recanto, e não imagina ver nenhuma ovelha. Mesmo assim, você não está confiante o suficiente, então olha para dentro do curral e parece haver muitas ovelhas ali. Você repassa seus esforços anteriores e decide que foi especialmente diligente. Isso alivia sua ansiedade e você dormirá. Na manhã seguinte, você descobre que duas ovelhas estão mortas. Algo dentro de você se quebra e você começa a pensar criativamente.

Naquele dia, barulhos altos de marteladas vêm do portão do curral.

Na manhã seguinte, você abre apenas um pouco o portão do curral e, à medida que cada ovelha sai, joga uma pedrinha em um balde pregado ao lado da porta. À tarde, quando cada ovelha retorna, você retira uma pedrinha do balde. Quando não houver mais pedrinhas no balde, você pode parar de procurar e dormir. É uma ideia brilhante que revolucionará o pastoreio.

Essa era a teoria. Na prática, foi necessário um refinamento considerável antes que o método funcionasse de forma confiável. Várias vezes procurei por horas e não encontrei nenhuma ovelha, e na manhã seguinte não havia nenhum desgarrado. Em cada uma dessas ocasiões, foi necessário pensar profundamente para descobrir onde meu sistema de balde falhou. Ao retornar de uma busca infrutífera, pensei no passado e percebi que o balde já continha pedrinhas quando comecei; isso acabou sendo uma má ideia. Outra vez, joguei pedrinhas ao acaso no balde, para me divertir, entre a manhã e a tarde; isso também foi uma má ideia, como percebi após procurar por algumas horas. Mas pratiquei meu ofício de seixo e me tornei um artesão de seixo razoavelmente proficiente.

Certa tarde, um homem ricamente vestido com túnicas brancas, louros frondosos, sandálias e um terno de negócios caminha penosamente ao longo da trilha arenosa que leva às minhas pastagens.

“Posso ajudar?” pergunto.

O homem tira um distintivo de seu casaco e o abre, provando sem sombra de dúvida que ele é Markos Sophisticus Maximus, um delegado do Senado de Rum. (Alguém pode se perguntar se outro poderia roubar o distintivo; mas é tão grande o poder desses distintivos que, se qualquer outro os usasse, eles seriam naquele instante transformados em Markos.)

“Me chame de Mark”, diz ele. “Estou aqui para confiscar as pedrinhas mágicas, em nome do Senado; artefatos de tão grande poder não devem cair em mãos ignorantes.”

“Aquele maldito aprendiz,” murmuro baixinho, “ele está tagarelando com os aldeões de novo.” Então, olho para o rosto severo de Mark e suspiro. “Não são pedrinhas mágicas”, digo em voz alta. “Apenas pedras comuns que peguei do chão.”

Um lampejo de confusão cruza o rosto de Mark, e então ele se ilumina novamente. “Estou aqui pelo balde mágico!” ele declara.

“Não é um balde mágico,” digo cansadamente. “Eu costumava guardar meias sujas nele.” O rosto de Mark está confuso. “Então, onde está a mágica?” Ele exige.

Uma pergunta interessante. “É difícil de explicar”, digo.

Meu atual aprendiz, Autrey, atraído pela comoção, se aproxima e oferece sua explicação: “É o nível de pedrinhas no balde”, diz Autrey. “Existe um nível mágico de pedrinhas, e você tem que acertar o nível, ou não funciona. Se você jogar mais pedrinhas ou tirar algumas, o balde não estará mais no nível mágico. No momento, o nível mágico está”, Autrey espiou dentro do balde, “cerca de um terço cheio”.

“Entendi!” Mark disse animado. Do bolso de trás, Mark tirou seu próprio balde e uma pilha de pedrinhas. Então, ele pegou alguns punhados de pedrinhas e as colocou no balde. Depois, Mark olhou para dentro do balde, observando quantas pedras havia lá. “Lá vamos nós”, disse Mark, “o nível mágico deste balde está meio cheio. Assim?”

“Não.” Autrey disse bruscamente. “Meio cheio não é o nível mágico. O nível mágico é de cerca de um terço. Meio cheio definitivamente não é mágico. Além disso, você está usando o balde errado.”

Mark se virou para mim, intrigado. “Eu pensei que você disse que o balde não era mágico?” “Não é,” eu disse. Uma ovelha passou pelo portão e eu joguei outra pedrinha no balde. “Além disso, estou cuidando das ovelhas. Fale com Autrey.”

Mark olhou duvidosamente para a pedra que joguei, mas decidiu arquivar temporariamente a pergunta. Mark se virou para Autrey e se levantou com altivez. “É um país livre”, disse Mark, “sob a benevolente ditadura do Senado, é claro. Posso jogar as pedrinhas que quiser no balde que quiser.”

Autrey considerou isso. “Não, você não pode,” ele disse finalmente, “não haverá nenhuma mágica.”

“Olhe”, disse Mark pacientemente, “observei você com cuidado. Você olhou dentro do seu balde, verificou o nível das pedras e chamou isso de nível mágico. Fiz o mesmo.”

“Não é assim que funciona”, disse Autrey.

“Ah, entendo”, disse Mark, “não é o nível de pedrinhas no meu balde que é mágico, é o nível de pedrinhas no seu balde. É isso que você afirma? O que torna o seu balde muito melhor do que o meu, hein?”

“Bem”, disse Autrey, “se esvaziássemos o seu balde e despejássemos todas as pedras do meu balde no seu, o seu balde teria o mesmo nível de pedras que o meu. Existe um procedimento que podemos usar para verificar se o seu balde tem o nível mágico, caso saibamos que o meu balde tem o nível mágico; chamamos isso de operação de comparação de balde.”

Outra ovelha passa e eu jogo outra pedrinha.

“Ele acabou de jogar outra pedrinha!”, disse Mark. “E suponho que você afirme que o novo nível também é mágico? Eu poderia jogar pedrinhas em seu balde até que o nível fosse o mesmo que o meu, e então nossos baldes concordariam. Você está apenas comparando meu balde com o seu para determinar se acha que o nível é “mágico” ou não. Bem, acho que seu balde não é mágico, porque não tem o mesmo nível de pedrinhas que o meu. Então é isso!”

“Espere”, disse Autrey, “você não entende…”

“Quando você diz ‘nível mágico’, você se refere simplesmente ao nível de pedras em seu próprio balde. E quando digo ‘nível mágico’, quero dizer o nível de pedras no meu balde. Assim, você olha para o meu balde e diz que ‘não é mágico’, mas a palavra ‘mágico’ significa coisas diferentes para pessoas diferentes. Você precisa especificar de quem é a magia. Você deve dizer que meu balde não tem o ‘nível mágico de Autrey’, e eu digo que seu balde não tem o ‘nível mágico de Mark’. Assim, a aparente contradição desaparece.”

“Mas…”, disse Autrey impotente.

“Pessoas diferentes podem ter baldes diferentes com níveis diferentes de pedras, o que prova que essa coisa de ‘mágica’ é completamente arbitrária e subjetiva.”

“Mark”, disse eu, “alguém já lhe disse o que essas pedras fazem?”

“Fazem?” disse Markos. “Eu pensei que elas eram apenas mágicas.”

“Se as pedras não fizessem nada”, disse Autrey, “nosso auditor de eficiência de processo ISO 9000 eliminaria o procedimento do nosso trabalho diário.”

“E qual é o nome do seu auditor?”, perguntei. “Darwin”, respondeu Autrey.”

“Hum”, disse Mark. “Charles tem a reputação de ser um auditor rigoroso. Então, os seixos abençoam os rebanhos e causam o aumento de ovelhas?”

“Não”, eu disse. “A virtude dos seixos é esta: se olharmos para o balde e vermos que o balde está vazio de seixos, sabemos que os pastos também estão vazios de ovelhas. Se não usarmos o balde, devemos procurar e procurar até o anoitecer, para não restar uma última ovelha. Ou, se pararmos nosso trabalho cedo, às vezes na manhã seguinte encontramos uma ovelha morta, pois os lobos atacam qualquer ovelha deixada do lado de fora. Se olharmos no balde, saberemos quando todas as ovelhas estão em casa e podemos nos aposentar sem medo”.

Mark considera isso. “Isso soa bastante implausível”, diz ele finalmente. “Você considerou usar varinhas de adivinhação? Varinhas de adivinhação são infalíveis ou, pelo menos, qualquer um que diga serem falíveis é queimado na fogueira. Esta é uma maneira extremamente dolorosa de morrer; segue-se que as varinhas de adivinhação são infalíveis.

“Você pode usar varinhas de adivinhação, se quiser”, eu disse.

“Oh, meu Deus, claro que não”, disse Mark. “Elas funcionam infalivelmente, com perfeição absoluta em todas as ocasiões, como convém a tais instrumentos abençoados; mas e se houvesse uma ovelha morta na manhã seguinte? Só uso varinhas de adivinhação quando não há possibilidade de provar que estão erradas. Caso contrário, posso ser queimado vivo. Então, como funciona o seu balde mágico?”

Como funciona o balde…? É melhor começar com o caso mais simples possível. “Bem”, eu disse, “suponha que os pastos estejam vazios e o balde não esteja vazio. Então vamos perder horas procurando uma ovelha que não está lá. E se houver ovelhas nos pastos, mas o balde estiver vazio, então Autrey e eu vamos nos recolher muito cedo e encontraremos ovelhas mortas na manhã seguinte. Portanto, um balde vazio é mágico se e somente se os pastos estiverem vazios…”

“Espere”, disse Autrey. “Isso soa como uma tautologia vazia para mim. Um balde vazio e pastos vazios não são obviamente o mesmo?”

“Não é vazio”, eu disse. “Aqui está uma analogia: o lógico Alfred Tarski disse uma vez que a afirmação ‘a neve é branca’ é verdadeira se e somente se a neve for branca. Se você puder entender isso, poderá ver por que um balde vazio é mágico se e somente se os pastos estiverem vazios de ovelhas.”

“Espere”, disse Mark. “São baldes. Eles não têm nada a ver com ovelhas. Baldes e ovelhas são obviamente completamente diferentes. Não há como as ovelhas interagirem com o balde.”

“Então de onde você acha que vem a mágica?” perguntou Autrey.

Mark considerou. “Você disse que poderia comparar dois baldes para verificar se eles tinham o mesmo nível… Posso ver como os baldes podem interagir com os baldes. Talvez quando você consegue uma grande coleção de baldes, e todos eles têm o mesmo nível, é isso que gera a mágica. Chamarei isso de teoria coerentista dos baldes mágicos.

“Interessante”, disse Autrey. “Eu sei que meu mestre está trabalhando em um sistema com vários baldes — ele diz que pode funcionar melhor por causa da ‘redundância’ e da ‘correção de erros’. Isso soa como coerentismo para mim.

“Eles não são os mesmos…” comecei a dizer.

“Vamos testar a teoria do coerentismo da magia”, disse Autrey. “Vejo que você tem mais cinco baldes no bolso de trás. Eu lhe darei o balde que estamos usando, e então você pode encher seus outros baldes até o mesmo nível…”

Mark recuou horrorizado. “Parar! Esses baldes são transmitidos na minha família há gerações e sempre tiveram o mesmo nível! Se eu aceitar seu balde, minha coleção de baldes ficará menos coerente e a magia desaparecerá!”

“Mas seus baldes atuais não têm nada a ver com as ovelhas!” protestou Autrey.

Mark parecia exasperado. “Olha, eu já expliquei antes, obviamente não há como as ovelhas interagirem com os baldes. Os baldes só podem interagir com outros baldes.”

“Eu jogo uma pedrinha sempre que uma ovelha passa”, ressaltei.

“Quando uma ovelha passa, você joga uma pedrinha?” Mark disse. “O que isso tem a ver com alguma coisa?”

“É uma interação da ovelha com as pedrinhas”, respondi.

“Não, é uma interação entre as pedras e você”, disse Mark. “A magia não vem das ovelhas, vem de você. Meras ovelhas obviamente não são mágicas. A mágica tem que vir de algum lugar, no caminho para o balde.”

Apontei para um mecanismo de madeira empoleirado no portão. “Você vê aquela aba de pano pendurada naquela engenhoca de madeira? Ainda estamos mexendo nisso — não funciona de forma confiável — mas quando as ovelhas passam, elas perturbam o tecido. Quando o pano se move para o lado, uma pedrinha cai de um reservatório e cai dentro do balde.

Mark franziu a testa. “Eu não estou te entendendo muito bem… o pano é mágico?” Dou de ombros. “Encomendei online de uma empresa chamada Natural Selections. O tecido é chamado de Modalidade Sensorial.” Faço uma pausa, vendo as expressões incrédulas de Mark e Autrey. “Admito que os nomes são um pouco new-age. A questão é que uma ovelha que passa desencadeia uma cadeia de causa e efeito que termina com uma pedrinha no balde. Depois, você pode comparar o balde com outros baldes e assim por diante.”

“Ainda não entendi”, disse Mark. “Você não pode colocar uma ovelha em um balde. Apenas seixos vão em baldes, e é óbvio que seixos só interagem com outros seixos.”

“As ovelhas interagem com coisas que interagem com pedrinhas…” Procuro uma analogia. “Imagine que você olhe para baixo nos seus cadarços. Um fóton deixa o Sol; em seguida, atravessa a atmosfera da Terra, salta dos seus cadarços, passa pela pupila do seu olho, atinge a retina, sendo absorvido por um bastonete ou cone. A energia do fóton faz com que o neurônio associado dispare, o que, por sua vez, faz com que outros neurônios disparem. Um padrão de ativação neural no seu córtex visual pode interagir com suas crenças sobre seus cadarços, já que crenças sobre cadarços também existem no substrato neural. Se você conseguir entender isso, poderá ver como uma ovelha que passa faz com que uma pedrinha entre no balde”.

“Em que ponto exato do processo a pedrinha se torna mágica?” disse Mark.

“Isso… hum…” Agora estou começando a ficar confuso. Balanço a cabeça para afastar as teias de aranha. Tudo isso parecia bastante simples quando acordei esta manhã, e o sistema de pedrinhas e baldes não ficou mais complicado desde então. “Isso é muito mais fácil de entender se você lembrar que o objetivo do sistema é rastrear as ovelhas.”

Mark suspirou tristemente. “Deixa pra lá… é óbvio que você não sabe. Talvez todas as pedrinhas sejam mágicas desde o início, mesmo antes de entrarem no balde. Podemos chamar essa posição de panpedrismo.

“Ha!” Autrey disse, desprezo evidente em sua voz. “Meras ilusões! Nem todas as pedras são criadas iguais. As pedrinhas do seu balde não são mágicas. São apenas pedaços de pedra!”

O rosto de Mark ficou sério. “Agora,” ele gritou, “você vê o perigo da estrada em que está andando! Após dizer que as pedras de algumas pessoas são mágicas e outras não, seu orgulho o consumirá! Você se achará superior a todos os outros e cairá! Muitos ao longo da história torturaram e assassinaram porque achavam que suas próprias pedras eram supremas!” Um tom de condescendência apareceu na voz de Mark. “Adorar um nível de seixos como ‘mágico’ implica que há um nível absoluto de seixo em um Balde Supremo. Ninguém acredita em um balde supremo atualmente.”

“Um”, eu disse. “Ovelhas não são pedras absolutas. Dois, não acredito que meu balde contenha realmente as ovelhas. Três, não considero meu nível de balde perfeito — às vezes, ajusto — e faço isso porque me preocupo com as ovelhas.

“Além disso”, disse Autrey, “alguém que acredita que possuir seixos absolutos permitiria tortura e assassinato está cometendo um erro que nada tem a ver com baldes. Você está resolvendo o problema errado.”

Mark se acalmou. “Acho que não posso esperar nada melhor de meros pastores. Você provavelmente acredita que a neve é branca, não é?

“Hum… Sim?” disse Autrey.

“Você não se incomoda que Joseph Stalin acreditasse que a neve é branca?” “Hum… não?”, disse Autrey.

Mark olhou incrédulo para Autrey e deu finalmente de ombros. “Vamos supor, apenas para fins de argumentação, que suas pedras são mágicas e as minhas não. Você pode me dizer qual é a diferença?”

“Minhas pedrinhas representam as ovelhas!” Autrey disse triunfante. “Suas pedrinhas não têm a propriedade de representatividade, então não vão funcionar. Eles são vazios de significado. Basta olhar para eles. Não há aura de conteúdo semântico; são apenas seixos. Você precisa de um balde com poderes causais especiais.

“Ah!” Mark disse. “Poderes causais especiais, em vez de magia.”

“Exatamente”, disse Autrey. “Não sou supersticioso. Postular magia, hoje em dia, seria inaceitável para a comunidade internacional de pastores. Descobrimos que postular magia simplesmente não funciona como uma explicação para fenômenos de pastoreio. Então, quando vejo algo que não entendo e quero explicá-lo usando um modelo sem detalhes internos que não faz previsões mesmo em retrospecto, eu postulo poderes causais especiais. Se isso não funcionar, passarei a chamá-lo de fenômeno emergente”.

“Que tipo de poderes especiais o balde tem?” perguntou Mark.

“Hum”, disse Autrey. “Talvez este balde esteja imbuído de uma relação de proximidade com os pastos. Isso explicaria por que funcionou — quando o balde está vazio, significa que os pastos estão vazios.

“Onde você encontrou esse balde?” disse Mark. “E como você percebeu ter uma relação de proximidade com os pastos?”

“É um balde comum”, eu disse. “Eu costumava subir em árvores com ele… Não acho que essa pergunta precise ser difícil.

“Estou falando com Autrey”, disse Mark.

“Você tem que amarrar o balde às pastagens e as pedras às ovelhas, usando um ritual mágico — perdoe-me, um processo emergente com poderes causais especiais — que meu mestre descobriu”, explicou Autrey.

Autrey então tentou descrever o ritual, com Mark concordando com a compreensão do sábio.

“Você tem que jogar uma pedrinha toda vez que uma ovelha sai pelo portão?” disse Mark. “Tirar uma pedrinha toda vez que uma ovelha voltar?”

Autrey concordou. “Sim.”

“Isso deve ser muito difícil”, disse Mark com simpatia.

Autrey se sentiu iluminada, absorvendo a simpatia de Mark como chuva. “Exatamente!”, disse Autrey. “É extremamente difícil controlar suas emoções. Quando o balde mantém seu nível por um tempo, você… tende a se apegar a esse nível”.

Em seguida, uma ovelha passou pelo portão. Autrey viu e se abaixou para pegar uma pedra, a ergue no ar e proclama: “Contemplem! Uma ovelha passou! Agora devo jogar uma pedrinha neste balde, meu querido balde, e destruir aquele nível afetuoso que durou tanto tempo… Outra ovelha passa. Autrey, envolvido em seu drama, erra o alvo e joga a pedra no balde. Autrey continua falando: “… pois esse é o teste supremo do pastor, jogar a pedra, por mais agonizante que seja, sendo o antigo nível sempre tão precioso. De fato, apenas o melhor dos pastores pode atender a uma exigência tão severa…”

“Autrey”, eu disse, “se você quer ser um grande pastor algum dia, aprenda a calar a boca e jogar a pedrinha. Sem confusões. Sem drama. Apenas faça.”

“E esse ritual”, disse Mark, “liga as pedras às ovelhas pelas leis mágicas da Simpatia e do Contágio, como uma boneca de vodu”.

Autrey estremeceu e olhou em volta. “Por favor! Não chame isso de simpatia e contágio. Nós, pastores, somos um povo antissupersticioso. Use a palavra ‘intencionalidade’ ou algo assim.”

“Posso olhar para uma pedra?” disse Mark.

“Claro”, eu disse. Peguei uma das pedrinhas do balde e joguei para Mark. Então, alcancei o chão, peguei outra pedrinha e joguei dentro do balde.

Autrey olhou para mim, intrigado. “Você não estragou tudo?”

Dei de ombros. “Eu não acho. Saberemos se estraguei tudo se houver uma ovelha morta na manhã seguinte ou se procurarmos por algumas horas e não encontrarmos nenhuma ovelha.

“Mas…” Autrey disse.

“Eu ensinei tudo o que você sabe, mas não ensinei tudo o que sabe,” eu disse.

Mark examinou a pedra, encarando-a atentamente. Ele segurou a mão sobre a pedra e murmurou algumas palavras, depois balança a cabeça. “Não sinto nenhum poder mágico”, disse ele. “Perdoe-me. Não sinto nenhuma intencionalidade.”

“Uma pedrinha só tem intencionalidade se estiver dentro de um balde ma… —um balde emergente”, disse Autrey. “Caso contrário, é apenas uma mera pedrinha.”

“Não há problema”, eu disse. Peguei uma pedrinha do balde e joguei fora. Então, fui até onde Mark estava, bati em sua mão que segurava uma pedrinha e disse: “Eu declaro que esta mão faz parte do balde mágico!” Então, retomo meu posto nos portões.

Autrey riu. “Agora você está apenas sendo maldoso gratuitamente.” Concordei com a cabeça, pois este era realmente o caso.

“Mas isso realmente vai funcionar?” disse Autrey.

Acenei de novo, esperando que estivesse certo. Já fiz isso antes com dois baldes e, em princípio, não deveria haver diferença entre a mão de Mark e um balde. Mesmo que a mão de Mark esteja imbuída do elã vital que distingue a matéria viva da matéria morta, o truque funcionaria tão bem quanto se Mark fosse uma estátua de mármore.

Mark estava olhando para sua mão, um pouco nervoso. “Então… a pedrinha tem intencionalidade de novo, agora?”

“Sim”, eu disse. “Não coloque mais pedrinhas em sua mão, nem jogue fora as que você tem, ou você quebrará o ritual.”

Mark acenou com a cabeça solenemente. Então, ele recomeçou a inspecionar a pedra. “Eu entendo agora como seus rebanhos cresceram tanto”, disse Mark. “Com o poder deste balde, você poderia continuar jogando pedrinhas, e as ovelhas continuariam voltando dos campos. Você poderia começar com apenas algumas ovelhas, deixá-las ir embora e depois encher o balde até a borda antes que elas voltassem. E se cuidar de tantas ovelhas se tornasse tedioso, você poderia deixá-las ir embora, depois esvaziar quase todas as pedras do balde, de modo que apenas algumas retornassem… aumentando os rebanhos novamente quando chegar a hora da tosquia… queridos céus, cara! Você percebe o poder absoluto deste ritual que você descobriu? Só posso imaginar as implicações; a humanidade pode saltar uma década à frente — não, um século!

“Não funciona assim,” eu disse. “Se você adicionar uma pedra quando uma ovelha não saiu ou remover uma pedra quando uma ovelha não entrou, isso quebra o ritual. O poder não permanece nas pedras, mas desaparece de uma só vez, como uma bolha de sabão estourando.”

O rosto de Mark ficou terrivelmente desapontado. “Tem certeza?” Eu concordei. “Eu tentei isso e não funcionou.”

Mark suspirou pesadamente. “E isto… matemática… parecia tão poderoso e útil até então… Ah, bem. Lá se vai o progresso humano.”

“Mark, foi uma ideia brilhante”, disse Autrey encorajadoramente. “A noção não me ocorreu, no entanto, é tão óbvia… isso economizaria uma quantidade enorme de esforço… deve haver uma maneira de salvar seu plano! Poderíamos experimentar diferentes baldes, procurando aquele que mantivesse o poder mágico — a intencionalidade nas pedrinhas, mesmo sem o ritual. Ou tente outras pedras. Talvez nossas pedras tenham apenas as propriedades erradas para ter uma intencionalidade inerente. E se tentássemos usar pedras esculpidas para se assemelhar a pequenas ovelhas? Ou apenas escreva ‘ovelha’ nas pedras; isso pode ser o suficiente.

“Não vai funcionar”, previ secamente.

Autrey continuou. “Talvez precisemos de seixos orgânicos, em vez de seixos de silício… ou talvez precisemos usar pedras preciosas caras. O preço das gemas dobra a cada dezoito meses, então você pode comprar um punhado de gemas baratas agora e esperar, e em vinte anos elas estarão muito caras.

“Você tentou adicionar pedrinhas para criar mais ovelhas e não funcionou?” Mark me perguntou. “O que exatamente você fez?”

“Peguei um punhado de notas de dólar. Então escondi as notas de dólar sob uma dobra do meu cobertor, uma a uma; cada vez que escondia outra nota, tirava outro clipe de papel de uma caixa, fazendo uma pequena pilha. Tomei o cuidado de não acompanhar tudo na minha cabeça, de modo que tudo o que sabia era que havia “muitas” notas de dólar e “muitos” clipes de papel. Então, quando todas as notas estavam escondidas sob meu cobertor, acrescentei um único clipe de papel à pilha, o equivalente a jogar uma pedrinha extra no balde. Então comecei a tirar notas de um dólar debaixo da dobra e colocar os clipes de volta na caixa. Quando terminei, sobrou um único clipe de papel.”

“O que esse resultado significa?” perguntou Autrey.

“Significa que o truque não funcionou. Assim que quebrei o ritual com aquele único passo em falso, o poder não durou, mas desapareceu instantaneamente; a pilha de clipes e a pilha de notas de dólar não ficavam mais vazias ao mesmo tempo.

“Você realmente tentou isso?” perguntou Mark.

“Sim,” eu disse, “realizei realmente o experimento, para verificar se o resultado correspondia à minha previsão teórica. Tenho um carinho sentimental pelo método científico, mesmo quando parece absurdo. Além disso, e se eu estivesse errado?

“Se tivesse funcionado”, disse Mark, “você seria culpado de falsificação!

Imagine se todos fizessem isso; a economia entraria em colapso! Todos teriam bilhões de dólares em moeda, mas não haveria nada para o dinheiro comprar!”

“De jeito nenhum”, respondi. “Pela mesma lógica em que adicionar outro clipe de papel à pilha cria outra nota de dólar, criar outra nota de dólar criaria um valor adicional em bens e serviços em dólares.”

Mark balançou a cabeça. “A falsificação ainda é crime. . . Você não deveria ter tentado.

“Eu estava razoavelmente confiante de que iria falhar.”

“Ahá!” disse Mark. “Você esperava falhar! Você não acreditou que poderia fazer isso!”

“De fato,” eu admiti. “Você adivinhou minhas expectativas com uma precisão impressionante.”

“Bem, esse é o problema”, disse Mark rapidamente. “A magia é alimentada pela crença e força de vontade. Se você não acredita que pode fazer isso, você não pode. Você precisa mudar sua crença sobre o resultado experimental; isso mudará o próprio resultado.”

“Engraçado”, eu disse nostalgicamente, “foi o que Autrey disse quando contei a ele sobre o método da pedrinha e do balde. Que era ridículo demais para ele acreditar, então não funcionaria para ele.

“Como você o persuadiu?” indagou Mark.

“Eu disse a ele para calar a boca e seguir as instruções”, eu disse, “e quando o método funcionou, Autrey começou a acreditar nele.”

Mark franziu a testa, confuso. “Isso não faz sentido. Não resolve o dilema essencial do ovo e da galinha.”

“Claro que sim. O método do balde funciona, quer você acredite nele ou não.”

“Isso é um absurdo!” reclamou Mark. “Eu não acredito em mágica que funciona, quer você acredite nela ou não!”

“Eu disse isso também”, acrescentou Autrey. “Aparentemente, eu estava errado.”

Mark franziu o rosto em concentração. “Mas… se você não acredita em mágica que funciona, quer você acredite ou não, então por que o método do balde funcionou quando você não acreditou nele? Você acreditou em magia que funciona, quer você acredite ou não, acredite ou não em magia que funciona, acredite ou não?”

“Eu não acho que é assim. .” disse Autrey em dúvida.

“Então, se você não acredita em mágica que funciona, quer você ou não… espere um segundo, preciso resolver isso com papel e lápis… Mark rabiscou freneticamente, olhou com ceticismo para o resultado, virou o pedaço de papel de cabeça para baixo e desistiu. “Não importa”, disse Mark. “A magia é difícil o suficiente para eu compreender; metamagia está fora do meu alcance.”

“Mark, acho que você não entende a arte da caçamba,” eu disse. “Não se trata de usar pedrinhas para controlar ovelhas. Trata-se de fazer pedrinhas de controle de ovelhas. Nesta arte, não é necessário começar acreditando que a arte funcionará. Em vez disso, primeiro a arte funciona, depois passa-se a acreditar que funciona.”

“Ou então você acredita”, disse Mark.

“Então eu acredito,” eu respondi, “porque acontece de ser um fato. A correspondência entre a realidade e minhas crenças vem da realidade controlando minhas crenças, e não o contrário.”

Outra ovelha passou, fazendo-me jogar outra pedra.

“Ah! Agora chegamos à raiz do problema”, disse Mark. “O que é esse negócio de ‘realidade’? Entendo o que significa uma hipótese ser elegante, ou falsificável, ou compatível com a evidência. Parece-me que chamar uma crença de ‘verdadeira’ ou ‘real’, ou ‘genuína’ é apenas a diferença entre dizer que você acredita em algo e dizer que realmente acredita em algo.

Eu pausei. “Bem….” e disse lentamente. “Francamente, eu mesmo não tenho muita certeza de onde vem esse negócio de ‘realidade’. Não posso criar minha própria realidade no laboratório, então não devo entendê-la ainda. Mas, ocasionalmente, acredito fortemente que algo acontecerá e, em vez disso, outra coisa acontece. Preciso de um nome para o que quer que seja que determina meus resultados experimentais, então chamo de ‘realidade’. Essa “realidade” é de alguma forma separada até mesmo das minhas melhores hipóteses. Mesmo quando tenho uma hipótese simples, fortemente apoiada por todas as evidências que conheço, às vezes ainda me surpreendo. Portanto, preciso de nomes diferentes para as coisas que determinam minhas previsões e para as coisas que determinam meus resultados experimentais. Chamo as primeiras coisas de ‘crença’ e as últimas de ‘realidade’.”

Mark bufou. “Eu nem sei por que me incomodo em ouvir esse absurdo óbvio. O que quer que você diga sobre essa chamada “realidade”, é apenas outra crença. Mesmo sua crença de que a realidade precede suas crenças é uma crença. Segue-se, como uma inevitabilidade lógica, que a realidade não existe; só as crenças existem.”

“Espere”, disse Autrey, “você poderia repetir a última parte? Você me perdeu ali no meio.

“Não importa o que você diga sobre a realidade, é apenas outra crença”, explica Mark. “Segue-se com esmagadora necessidade que não há realidade, apenas crenças.” “Entendo”, eu disse. “Da mesma forma que não importa o que você coma, você precisa comer com a boca. Segue-se que não há comida, apenas bocas.”

“Exatamente”, disse Mark. “Tudo o que você come tem que estar na sua boca. Como pode haver comida que existe fora de sua boca? O pensamento é absurdo, provando que ‘comida’ é uma noção incoerente. É por isso que todos estamos morrendo de fome; não tem comida”.

Autrey olhou para sua barriga. “Mas não estou morrendo de fome.”

“Ahá!” gritou Mark triunfantemente. “E como você expressou essa mesma objeção? Com a boca, meu amigo! Com a boca! Que melhor demonstração você poderia pedir de que não há comida?”

“A voz áspera e rouca que vinha diretamente atrás de nós, perguntou: “O que é isso sobre fome?” Autrey e eu mantivemos a calma, já que havíamos passado por isso antes. Mark deu um salto de susto, quase fora de si.”

O inspetor Darwin sorriu amplamente, contente com a surpresa, e fez um pequeno sinal em sua prancheta.

“Apenas uma metáfora!” Mark disse rapidamente. “Você não precisa tirar minha boca, ou qualquer coisa assim.”

“Por que você precisa de uma boca se não há comida?” Darwin exige com raiva. “Deixa pra lá. Não tenho tempo para essa tolice. Estou aqui para inspecionar as ovelhas”.

“O rebanho está prosperando, senhor”, eu disse. “Nenhuma ovelha morreu desde janeiro”.

“Excelente. Eu o recompensarei com 0,12 unidades de condicionamento físico. Agora, o que essa pessoa está fazendo aqui? Ele é uma parte necessária das operações?

“Até onde posso ver, ele seria mais útil para a espécie humana se fosse pendurado em um balão de ar quente como lastro”, eu disse.

“Ai”, disse Autrey suavemente.

“Eu não me importo com a espécie humana. Deixe-o falar por si mesmo.”

Mark levantou-se com altivez. “Este mero pastor”, disse ele, gesticulando para mim, “alegou que existe algo chamado realidade. Isso me ofende, pois sei com profunda e permanente certeza que não há verdade. O conceito de ‘verdade’ é apenas um estratagema para as pessoas imporem suas próprias crenças aos outros. Cada cultura tem uma ‘verdade’ diferente, e a ‘verdade’ de nenhuma cultura é superior à outra. Isso que eu disse vale em todos os lugares, o tempo todo, e insisto que você concorde”.

“Espere um segundo”, disse Autrey. “Se nada é verdade, por que eu deveria acreditar em você quando você diz que nada é verdade?”

“Eu não disse que nada é verdade…”, disse Mark. “Sim, você disse”, interrompeu Autrey, “eu ouvi você”.

“Eu disse que ‘verdade’ é uma desculpa usada por algumas culturas para impor suas crenças a outras. Então, quando você diz que algo é ‘verdadeiro’, você quer dizer apenas que seria vantajoso para o seu próprio grupo social que acreditasse nisso”.

“E isso que você disse”, disse eu, “é verdade?”

“Absolutamente, positivamente verdade!”, disse Mark enfaticamente. “As pessoas criam suas próprias realidades”.

“Espere”, disse Autrey, parecendo confuso novamente, “dizer que as pessoas criam suas próprias realidades é, logicamente, uma questão completamente diferente de dizer que não há verdade, um estado de coisas que nem consigo imaginar coerentemente, talvez porque você ainda não explicou exatamente como isso deveria funcionar…”

“Lá vai você de novo”, disse Mark exasperado, “tentando aplicar seus conceitos ocidentais de lógica, racionalidade, razão, coerência e autoconsistência”.

“Ótimo”, murmurou Autrey, “agora preciso adicionar um terceiro cabeçalho de assunto, para acompanhar essa reivindicação totalmente separada e distinta…”

“Não é separado”, disse Mark. “Olha, você está abordando isso da maneira errada ao considerar minhas declarações como hipóteses e derivar cuidadosamente suas consequências. Você precisa pensar nelas

como desculpas totalmente gerais, que eu aplico quando alguém diz algo que eu não gosto. Não é tanto um modelo de como o universo funciona, mas sim um cartão "Saia da prisão sem pagar nada". A chave é aplicar a desculpa seletivamente. Quando digo não haver verdade, isso se aplica apenas à sua afirmação de que o balde mágico funciona, independentemente de eu acreditar nele ou não. Não se aplica à minha afirmação de que não há verdade".

"Hmm… por que não?" perguntou Autrey.

Mark suspirou pacientemente. "Autrey, você acha que é a primeira pessoa a pensar nessa pergunta? Perguntar como nossas próprias crenças podem ser significativas se todas as crenças são sem sentido? Isso é o mesmo que muitos estudantes dizem quando se deparam com essa filosofia, que, saibam, tem muitos seguidores e uma extensa literatura."

"Então, qual é a resposta?" perguntou Autrey.

"Chamamos isso de 'problema de reflexividade'", explicou Mark. "Mas qual é a resposta real?", persistiu Autrey.

Mark sorriu condescendentemente. "Acredite em mim, Autrey, você não é a primeira pessoa a pensar em uma pergunta tão simples. Não faz sentido apresentá-lo a nós como uma refutação triunfante."

"Mas qual é a resposta mesmo?"

"Agora, eu gostaria de abordar a questão de como a lógica mata focas fofinhas…", disse Mark.

"Você está perdendo tempo", retrucou o inspetor Darwin.

"Sem falar em perder o rastro das ovelhas", eu disse, jogando outra pedra.

O inspetor Darwin olhou para os dois argumentadores, ambos aparentemente relutantes em desistir de suas posições. "Ouça", disse Darwin, agora com mais gentileza, "tenho uma ideia simples para resolver a sua disputa. Você diz", disse Darwin, apontando para Mark, "que as crenças das pessoas alteram suas realidades pessoais. E você, fervorosamente, acredita", seu dedo girou para apontar para Autrey, "que as crenças de Mark não podem alterar a realidade. Então, deixe Mark acreditar fervorosamente que ele pode voar e, em seguida, caia de um penhasco. Mark se verá voando como um pássaro, e Autrey o verá caindo e se espatifando, e vocês dois ficarão felizes."

Todos fizemos uma pausa, considerando isso.

"Parece razoável…", disse Mark finalmente.

"Há um penhasco bem ali", observou o inspetor Darwin.

Autrey estava com um olhar de intensa concentração. Finalmente, ele gritou: "Espera! Se isso fosse verdade, todos teríamos partido há muito tempo para nossos próprios universos particulares, caso em que as outras pessoas aqui seriam apenas produtos da nossa imaginação. Não faz sentido tentar provar nada para nós…"

Um longo e minguante grito vem do penhasco próximo, seguido por um monótono e solitário som de impacto. O inspetor Darwin vira sua prancheta para a página que mostra o pool genético atual e escreve uma frequência um pouco menor para os alelos de Mark.

Autrey parece um pouco doente. "Aquilo era mesmo necessário?"

"Necessário?" diz o inspetor Darwin, parecendo confuso. "Acabou de acontecer…

Não entendi muito bem sua pergunta."

Autrey e eu voltamos para o nosso balde. É hora de trazer as ovelhas. Não queremos esquecer essa parte. Caso contrário, qual seria o sentido?"