LIVRO 4

MERA REALIDADE

RACIONALIDADE

De A a Z

ELIEZER YUDKOWSKY

RACIONALIDADE DE A a Z

MERA REALIDADE

LIVRO 4

por **ELIEZER YUDKOWSKY**

Tradução de Mariana Hungria
Revisão de Enéas Canavezzi Versehgi

Brasil, 2024

Sumário

O Mundo: uma introdução

por Rob Besinger

Em ensaios anteriores, discuti o raciocínio humano, a linguagem, os objetivos e a dinâmica social. Citei a matemática, a física e a biologia para explicar os padrões do comportamento humano, mas falei pouco sobre o lugar da humanidade na natureza ou sobre o próprio mundo natural.

Assim como foi útil contrastar os humanos como sistemas orientados a objetivos em processos humanos com a biologia evolutiva e a inteligência artificial, será útil nas próximas sequências de ensaios contrastar os humanos como sistemas físicos com processos inumanos que não têm nada a ver com a mente.

Afinal, nós, humanos, somos feitos de partes inumanas. O mundo dos átomos não se parece em nada com o mundo como normalmente o percebemos e certamente não se parece em nada com os habitantes conscientes do mundo como normalmente os concebemos. Como Giulio Giorello disse em uma entrevista com Daniel Dennett: "Sim, nós temos uma alma. Mas ela é feita de vários robôs minúsculos." [1]

"Mera realidade" reúne sete sequências de ensaios sobre esse tópico. As três primeiras introduzem a questão de como o mundo humano se relaciona com o mundo revelado pela física: "Verdade Legal" (sobre as conexões básicas entre física e cognição humana), "Reducionismo 101" (sobre o projeto de explicar fenômenos cientificamente) e "Alegria no Meramente Real" (sobre o significado emocional e pessoal da visão científica do mundo). Seguem-se duas sequências que se aprofundam em debates acadêmicos específicos: "Fisicalismo 201" (sobre o problema difícil da consciência) e "Física Quântica e Muitos Mundos" (sobre o problema da medição em física). Por fim, a sequência "Ciência e Racionalidade" e o ensaio "Uma Explicação Técnica da Explicação Técnica" conectam essas ideias e as relacionam com a prática científica.

As discussões sobre consciência e física quântica ilustram a relevância do reducionismo para as controvérsias atuais na ciência e na filosofia. Para os interessados em um contexto extra, direi mais algumas palavras sobre esses dois tópicos aqui. Para aqueles ansiosos para avançar: pule adiante!

Mentes no mundo: podemos saber como é ser um morcego?

Certamente podemos desenvolver modelos cognitivos mais sofisticados para prever o comportamento dos morcegos ou modelos mais refinados da neurologia dos morcegos, mas não é óbvio que isso nos diria qual é a sensação subjetiva da ecolocalização, ou da sensação de voar, do ponto de vista do morcego.

Na verdade, parece que nunca poderíamos ter certeza de que realmente existe algo como ser um morcego. Por que um autômato inconsciente não poderia replicar todos os comportamentos observáveis de um agente consciente com precisão arbitrária? (Os filósofos chamam esses autômatos de "zumbis", embora eles não tenham nada a ver com os zumbis do folclore — os quais são claramente diferentes dos agentes conscientes!)

Uma raça de psicólogos alienígenas se depararia com o mesmo problema se tentasse modelar a consciência humana. Eles poderiam chegar a um modelo preditivo perfeito do que dizemos e fazemos quando vemos uma rosa-vermelha, mas isso não significaria que os alienígenas compreendem completamente como é a vermelhidão subjetivamente.

Com exemplos como esses, filósofos como Thomas Nagel e David Chalmers argumentam que mo-

delos cognitivos e neurais de terceira pessoa nunca podem capturar totalmente a consciência de primeira pessoa. [2] [3] Não importa o quanto saibamos sobre um sistema físico, é sempre logicamente possível que o sistema não tenha experiências em primeira pessoa. O dualismo tradicional, com suas almas imateriais flutuando e violando livremente as leis físicas, pode ser falso; mas Chalmers defende uma tese mais fraca, a de que a consciência é um "fato adicional" não totalmente explicável pelos fatos físicos.

Vários filósofos e cientistas acharam essa linha de raciocínio persuasiva.[4] Se sentirmos a força intuitiva desse argumento, deveríamos admitir sua conclusão e abandonar o fisicalismo?

Certamente não devemos rejeitá-lo apenas porque soa estranho ou parece vagamente não científico. Mas como o argumento se sustenta diante de uma compreensão técnica de como a explicação e a crença funcionam? Existem algumas pistas obteníveis da história da ciência ou de nossa compreensão dos mecanismos físicos subjacentes às evidências? "Fisicalismo 201" voltará a esta questão.

Mundos no mundo

A mecânica quântica é o nosso melhor modelo matemático do universo até o momento, amplamente validado por um século de testes. A teoria postula uma "amplitude de probabilidade" de números complexos, assim chamada porque uma operação específica (elevar ao quadrado o valor absoluto do número - a regra de Born[1]), nos permite prever probabilisticamente fenômenos em pequenas escalas e níveis extremos de energia. Essa amplitude muda de forma determinística conforme a equação de Schrödinger[2]. No processo, ela frequentemente entra em estados chamados de "superposições".

No entanto, quando realizamos experimentos, as superposições quânticas parecem desaparecer sem deixar vestígios. Enquanto não estamos observando, a equação de Schrödinger parece capturar tudo sobre a dinâmica dos sistemas físicos. Mas quando observamos, esse determinismo é substituído pela regra probabilística de Born. É como se as leis ordinárias da física fossem repentinamente suspensas sempre que fazemos "observações". Como John Stewart Bell colocou:

> Parece que a teoria se preocupa exclusivamente com os "resultados das medições" e não oferece nenhuma informação sobre qualquer outra coisa. O que exatamente qualifica certos sistemas físicos para desempenhar o papel de "medidor"? A função de onda do mundo estava esperando para saltar por milhares de milhões de anos até que uma criatura viva unicelular aparecesse? Ou teria que esperar um pouco mais, por algum sistema mais qualificado… com um doutorado?[3]

Todos concordam que essa estranha combinação das regras de Schrödinger e de Born provou ser empiricamente adequada. Entretanto, a questão de saber exatamente quando a regra de Born entra em ação e o que isso significa causou um caos de diferentes pontos de vista sobre a natureza da mecânica quântica.

No início, a escola de Copenhague — Niels Bohr e outros criadores da teoria quântica — dividiu-se em várias formas padronizadas de falar sobre os resultados experimentais e o estranho formalismo usado para prevê-los. Alguns, levando o foco da teoria em "medidas" e "observações" literalmente, propuseram

---

1 NT. A **Regra de Born**, proposta por Max Born em 1926, é um postulado fundamental da mecânica quântica que estabelece como determinar a probabilidade de obter um determinado resultado ao medir um sistema físico. Segundo a regra, se o estado quântico de um sistema é descrito por uma função de onda ψ, a probabilidade de detectar o sistema em um estado específico (ou com um valor específico de uma grandeza física) é proporcional ao quadrado do módulo da amplitude de probabilidade associada a esse estado, ou seja, $|\psi|^2$. Essa regra conecta o formalismo matemático abstrato da teoria com resultados experimentais observáveis.

2 NT. A **equação de Schrödinger**, proposta por Erwin Schrödinger em 1925, é o pilar da mecânica quântica. Ela descreve como o estado quântico (função de onda) de um sistema físico evolui ao longo do tempo, permitindo prever probabilisticamente as propriedades observáveis das partículas.

3 NT. Texto original em inglês. *It would seem that the theory is exclusively concerned about "results of measurements" and has nothing to say about anything else. What exactly qualifies some physical systems to play the role of the "measurer"? Was the wavefunction of the world waiting to jump for thousands of millions of years until a single-celled living creature appeared? Or did it have to wait a little longer, for some better qualified system . . . with a PhD?*

que a consciência desempenhasse um papel fundamental na lei física, intervindo para fazer com que amplitudes complexas "colapsem" em observáveis. Outros, liderados por Werner Heisenberg, defendiam uma visão não realista segundo a qual a física trata de nossos estados de conhecimento e não de qualquer realidade objetiva. Ainda outra tradição de Copenhague, resumida no slogan "cale a boca e calcule", alertou contra especulações metafísicas de todos os tipos.

Yudkowsky usa essa controvérsia científica como campo de prova para algumas ideias centrais das sequências anteriores: distinções mapa-território, respostas misteriosas, Bayesianismo e Navalha de Ocam. Como ele não é físico — e nem eu — fornecerei algumas fontes externas aqui para os leitores que desejam examinar seus argumentos ou aprender mais sobre seus exemplos de física.

O livro *Our Mathematical Universe* (Nosso universo matemático) de Tegmark discute uma série de ideias relevantes em filosofia e física. [5] Entre as ideias mais inovadoras de Tegmark está seu argumento de que todas as estruturas matemáticas consistentes existem, incluindo mundos com leis físicas e condições iniciais totalmente diferentes das nossas. Ele distingue esses mundos de Tegmark dos multiversos em hipóteses cientificamente mais convencionais - por exemplo, mundos em modelos inflacionários estocásticos eternos do Big Bang e a interpretação de muitos mundos de Hugh Everett da física quântica.

Yudkowsky discute interpretações de muitos mundos em mais detalhes, principalmente como uma resposta às interpretações de Copenhague da mecânica quântica. Nas últimas décadas, os muitos mundos se tornaram muito populares entre os físicos, especialmente entre os cosmologistas. Entretanto, vários físicos ainda a rejeitam ou mantêm uma posição agnóstica. Para uma introdução filosófica ao debate, consulte o livro *Quantum Mechanics and Experience* (Mecânica Quântica e Experiência) de David Z. Albert. [6] A Stanford Encyclopedia of Philosophy também oferece introduções valiosas em seus artigos *Measurement in Quantum Theory* (Medição em Teoria Quântica) [7], *Everett's Relative-State Formulation* (A formulação do estado-relativo de Everett) [8] e *Many-Worlds Interpretation* (Interpretação de muitos mundos) [9].

No lado menos teórico, o livro *Thinking Physics* (Pensando a física) de Epstein é um ótimo texto para treinar intuições físicas. [10] Vale a pena ter em mente que, assim como se pode compreender a maior parte da ciência cognitiva sem compreender a natureza da consciência subjetiva, pode-se compreender a maior parte da física sem ter uma visão estabelecida da natureza definitiva (e do tamanho!) do mundo físico.

Referências

[1] Daniel C. Dennett, Freedom Evolves (Viking Books, 2003)

[2] David J. Chalmers, The Conscious Mind: In Search of a Fundamental Theory (New York: Oxford University Press, 1996).

[3] Thomas Nagel, “What Is It Like to Be a Bat?” Philosophical Review 83, no. 4 (1974): 435–450, http://www.jstor.org/stable/2183914

[4] Em uma pesquisa com filósofos profissionais anglófonos, 56,5% endossaram o fisicalismo, 27,1% endossaram o antifísico e 16,4% endossaram outros pontos de vista (por exemplo, “não sei”). [11] A maioria dos filósofos rejeita a possibilidade metafísica dos “zumbis” de Chalmers, mas não há consenso sobre o porquê, exatamente, o argumento dos zumbis de Chalmers falha. Kirk resume as posições contemporâneas sobre a consciência fenomenal, apresentando argumentos que se assemelham aos de Yudkowsky contra a possibilidade de conhecer ou referir-se a qualia irredutíveis. [12]

[5] Max Tegmark, Our Mathematical Universe: My Quest for the Ultimate Nature of Reality (Random House LLC, 2014).

[6] David Z. Albert, Quantum Mechanics and Experience (Harvard University Press, 1994).

[7] Henry Krips, “Measurement in Quantum Theory”, em The Stanford Encyclopedia of Philosophy, Outono de 2013, ed. Edward N. Zalta.

[8] Jeffrey Barrett, Everett’s Relative-State Formulation of Quantum Mechanics, ed. Edward N. Zalta,http://plato.stanford.edu/archives/fall2008/entries/qm-everett/

[9] Lev Vaidman, “Many-Worlds Interpretation of Quantum Mechanics”, em The Stanford Encyclopedia of Philosophy, outono de 2008, ed. Edward N. Zalta. Edward N. Zalta.

[10] Lewis Carroll Epstein, Thinking Physics: Understandable Practical Reality, 3ª edição (Insight Press, 2009).

[11] David Bourget e David J. Chalmers, “What Do Philosophers Believe?”, Philosophical Studies (2013): 1-36.

[12] Robert Kirk, Mind and Body (McGill-Queen’s University Press, 2003).

Parte O — Verdade Legal

181 — Fogo universal

Na história de fantasia de L. Sprague de Camp, *The Incomplete Enchanter* (O encantador incompleto) (que estabeleceu o modelo para as muitas imitações que se seguiram), o herói, Harold Shea, é transportado do nosso universo para o universo da mitologia nórdica. [1] Este mundo é baseado em magia e não em tecnologia; então, naturalmente, quando o nosso herói tenta acender uma fogueira com um fósforo trazido da Terra, o fósforo não pega fogo.

Sei que era apenas uma história de fantasia, mas… como posso dizer isso…

Não.

No final do século XVIII, Antoine-Laurent de Lavoisier descobriu o fogo. "Como assim?" você diz. "O uso do fogo não remonta a centenas de milhares de anos?" Bem, sim, as pessoas usavam o fogo; ele era quente, brilhante, meio alaranjado, e você podia usá-lo para cozinhar coisas. Mas ninguém sabia como ele funcionava. Os alquimistas gregos e medievais pensavam que o fogo era um elemento básico, um dos quatro elementos. Na época de Lavoisier, o paradigma alquímico havia sido gradualmente alterado e bastante complicado, mas o fogo ainda era considerado um elemento básico — na forma de "flogisto", uma substância bastante misteriosa que supostamente explicava o fogo e também todos os outros fenômenos da alquimia.

A grande inovação de Lavoisier foi pesar todas as peças do quebra-cabeça químico, tanto antes quanto depois da reação química. Antes de Lavoisier, acreditava-se que algumas transmutações químicas alteravam o peso total do material: se o antimônio finamente moído fosse exposto à luz solar focalizada por uma lente, após uma hora ele seria reduzido a cinzas, que pesariam um décimo a mais do que o antimônio original - mesmo que a combustão fosse acompanhada pela perda de uma espessa fumaça branca. No entanto, ao pesar todos os componentes das reações, incluindo o ar no qual elas ocorriam, Lavoisier descobriu que a matéria não era criada nem destruída. Se as cinzas queimadas aumentavam de peso, havia uma diminuição correspondente no peso do ar.

Lavoisier também sabia separar gases e descobriu que uma vela acesa diminuía a quantidade de um tipo de gás, o ar vital, e produzia outro gás, o ar fixo. Hoje nós os chamamos de oxigênio e dióxido de carbono. Quando o ar vital se esgotava, o fogo se apagava. Poderíamos supor, talvez, que a combustão transformasse o ar vital em ar fixo e o combustível em cinzas, e que a capacidade dessa transformação continuar fosse limitada pela quantidade de ar vital disponível.

A proposta de Lavoisier contradizia diretamente a então vigente teoria do flogisto. Só isso já seria chocante o suficiente, mas também aconteceu…

Para apreciar o que segue, você deve se colocar no estado de espírito do século XVIII. Esqueça a descoberta do DNA, que ocorreu apenas em 1953. Desaprenda a teoria celular da biologia, formulada em 1839. Imagine olhar para sua mão, flexionar os dedos… e não ter absolutamente nenhuma ideia de como ela funciona. A anatomia do músculo e do osso era conhecida, mas ninguém tinha noção de "o que faz isso funcionar" — por que um músculo se movia e se flexionava, enquanto a argila moldada em uma forma semelhante apenas ficava lá. Imagine seu próprio corpo composto de algo misterioso e incompreensível. E então, imagine descobrir…

… Que os humanos, ao respirar, consumiam o ar vital e expiravam o ar fixo. As pessoas também fa-

ziam combustão! Lavoisier mediu a quantidade de calor que os animais (e seu assistente Seguin) produziam durante o exercício, a quantidade de ar vital consumido e o ar fixo expirado. Quando os animais produziam mais calor, consumiam mais ar vital e exalavam mais ar fixo. As pessoas, como o fogo, consumiam combustível e oxigênio; as pessoas, como o fogo, produziam calor e dióxido de carbono. Prive as pessoas de oxigênio, ou de combustível, e a luz se apaga.

Os fósforos pegam fogo devido ao fósforo - os "fósforos de segurança" têm fósforo na faixa de ignição; os fósforos que acendem em qualquer lugar têm fósforo nas cabeças dos palitos. O fósforo é altamente reativo; o fósforo puro brilha no escuro e pode entrar em combustão espontânea. (Henning Brand, que purificou o fósforo em 1669, anunciou que havia descoberto o Fogo Elemental.) O fósforo também é adequado para seu papel no trifosfato de adenosina, ATP, o principal método do corpo para armazenar energia química. Às vezes, o ATP é chamado de "moeda molecular". Ele energiza seus músculos e ativa seus neurônios. Quase todas as reações metabólicas na biologia dependem do ATP e, portanto, das propriedades químicas do fósforo.

Se um fósforo parar de funcionar, você também para. Você não pode mudar apenas uma coisa.

As regras de nível superficial, "Fósforos pegam fogo quando riscados" e "Humanos precisam de ar para respirar", não estão obviamente conectadas. Demorou séculos para descobrir a conexão e, mesmo assim, ainda parece um fato distante aprendido na escola, relevante apenas para alguns especialistas. É muito fácil imaginar um mundo onde uma regra superficial é válida e a outra não; suprimir nossa credibilidade em uma crença, mas não na outra. Mas isso é imaginação, não realidade. Se o seu mapa se divide em quatro partes para facilitar o armazenamento, isso não significa que o território também está dividido em partes desconectadas. Nossas mentes armazenam diferentes regras superficiais em diferentes compartimentos, mas isso não reflete nenhuma divisão nas leis que regem a Natureza.

Podemos levar a lição adiante. O fósforo deriva seu comportamento de leis ainda mais profundas, eletrodinâmica e cromodinâmica. "Fósforo" é apenas nossa palavra para elétrons e quarks arranjados de uma certa maneira. Você não pode alterar as propriedades químicas do fósforo sem alterar as leis que regem os elétrons e os quarks.

Se você entrasse em um mundo onde os fósforos falhassem em acender, você deixaria de existir como matéria organizada.

A realidade está muito mais entrelaçada do que os humanos gostariam de acreditar.

Referências

[1] Lyon Sprague de Camp and Fletcher Pratt, Re Incomplete Enchanter (New York: Henry Holt & Company, 1941).

182 — Lei universal

Antoine-Laurent de Lavoisier descobriu que a ventilação (respiração) e o fogo (combustão) funcionavam pelo mesmo princípio. Foi uma das unificações mais surpreendentes da história da ciência, por reunir o reino mundano da matéria e o reino sagrado da vida, que até então os humanos dividiam em domínios separados.

A primeira grande simplificação foi a de Isaac Newton, que unificou o movimento dos planetas com a trajetória de uma maçã caindo. O impacto dessa descoberta foi muito maior do que o de Lavoisier. Não se tratava apenas do fato que Newton ousara unificar o reino terrestre da matéria comum com o reino celestial obviamente diferente e sagrado, outrora considerado a morada dos deuses. A descoberta de Newton deu origem à noção de uma lei universal, sendo a mesma em todos os lugares e em todos os tempos, sem nenhuma exceção.

Os seres humanos residem em um mundo de fenômenos superficiais, organizados em categorias com numerosas exceções. O comportamento de um tigre difere daquele de um búfalo. Enquanto a maioria dos búfalos possui quatro patas, alguns podem ter apenas três. Por conseguinte, por que alguém acreditaria que existem leis universais? É tão obviamente falso.

A única ocasião em que parece que queremos que uma lei seja válida em todos os lugares é quando estamos falando sobre leis morais — regras tribais de comportamento. Alguns membros da tribo podem tentar pegar mais do que sua parte justa da carne de búfalo — talvez inventando alguma desculpa esperta — então, no caso das leis morais, parece que temos um instinto para a universalidade. Sim, a regra de dividir a carne igualmente se aplica a você, agora mesmo, quer você goste ou não. Mas ainda assim há exceções. Se, por algum motivo bizarro, uma tribo mais poderosa ameaçasse matar todos vocês, a menos que Bob recebesse o dobro de carne somente nesta ocasião, você daria a Bob o dobro de carne. A concepção de uma regra sem nenhuma exceção, literalmente, parece insamente rígida, produto de um pensamento dogmático de extremistas tão presos à sua única ideia que não conseguem compreender a diversidade e a complexidade do universo real.

Essa é a acusação frequente contra os cientistas - os estudiosos profissionais da riqueza e da complexidade do universo real. Porque quando se observa de fato o universo, ele acaba sendo, pelos padrões humanos, insanamente rígido na aplicação de suas regras. Até onde sabemos, não houve uma única violação da Conservação do Momento desde o início dos tempos até agora.

Às vezes - muito raramente - observamos uma violação aparente de nossos modelos das leis fundamentais. Embora nossos modelos científicos possam durar uma ou duas gerações, eles não são estáveis ao longo dos séculos... mas não pense que isso implica que o universo seja caprichoso. Isso é confundir o mapa com o território. Pois quando a poeira abaixa e a teoria antiga é derrubada, descobrimos que o universo sempre agiu conforme a nova generalização que descobrimos, que, mais uma vez, é absolutamente universal até onde o conhecimento da humanidade se estende. Quando foi descoberto que a gravitação newtoniana é um caso especial da Relatividade Geral, viu-se que a Relatividade Geral governava a órbita de Mercúrio décadas antes que qualquer ser humano soubesse disso; e, mais tarde, ficaria evidente que a Relatividade Geral vinha governando o colapso das estrelas por bilhões de anos antes da humanidade. Apenas nosso modelo estava equivocado - a lei em si sempre foi absolutamente constante - ou assim diz nosso novo modelo.

Posso ter apenas 80% de confiança de que o limite da velocidade da luz durará os próximos cem mil

anos, mas isso não significa que acho que o limite da velocidade da luz é válido apenas 80% do tempo, com exceções ocasionais. A proposição à qual atribuo 80% de probabilidade é que a lei da velocidade da luz é absolutamente inviolável em todo o espaço e tempo.

Uma das razões pelas quais os gregos antigos não descobriram a ciência é que eles não perceberam que você poderia generalizar a partir de experimentos. Os filósofos gregos estavam interessados em fenômenos “normais”. Se você montar um experimento artificial, provavelmente obterá um resultado “monstruoso”, que não teria implicações sobre como as coisas realmente funcionam.

Portanto, é assim que os seres humanos tendem a sonhar, antes de aprenderem melhor; mas e os sonhos silenciosos do próprio universo que ele sonhou para si mesmo antes de sonhar com os seres humanos?

Se você quiser aprender a pensar como a realidade, aqui está o Tao:

Desde o início

nem uma única coisa incomum

jamais aconteceu.

183 — A realidade é feia?

Considere os cubos, {1, 8, 27, 64, 125, … }. Suas primeiras diferenças {7, 19, 37, 61, … } podem a princípio parecer não ter um padrão óbvio, mas ao calcular as segundas diferenças {12, 18, 24, …}, chegamos a um nível de relação mais simples. Ao calcular as terceiras diferenças {6, 6, … }, alcançamos o nível perfeitamente estável, onde o caos se dissolve em ordem.

Mas este é um exemplo escolhido a dedo. Talvez o "mundo real confuso" não tenha a beleza desses objetos matemáticos abstratos? Ou, talvez, fosse mais apropriado falar sobre neurociência ou redes de expressão gênica?

A matemática abstrata, sendo construída apenas na imaginação, surge de fundamentos simples - um pequeno conjunto de axiomas iniciais - e é um sistema fechado; condições que podem parecer anormalmente favoráveis à clareza.

Ou seja: na matemática pura, você não precisa se preocupar com um tigre saltando dos arbustos e comendo o Triângulo de Pascal.

Então o mundo real é mais feio do que a matemática?

É estranho que as pessoas ainda façam essa pergunta. A dúvida poderia ter sido razoável há dois milênios e meio, quando os filósofos gregos discutiam a composição do "mundo real", havia muitas perspectivas diferentes sobre o assunto. Heráclito disse: "Tudo é fogo". Tales disse: "Tudo é água". Pitágoras disse: "Tudo é número".

Pontuação:

Heráclito	0
Tales	0
Pitágoras	1

Sob as formas e os formatos complexos do mundo superficial, existe um nível simples, um nível exato e estável, cujas leis chamamos de "física". Essa descoberta, a Grande Surpresa, já ocorreu em nosso ponto da história humana — mas não devemos esquecer que ela foi surpreendente. Era uma vez, as pessoas procuravam a beleza subjacente sem garantia de encontrá-la; e era uma vez, elas a encontraram; e agora tornou-se uma coisa conhecida e dada como certa.

Então, por que não podemos prever a localização de cada tigre a espreita nos arbustos com a mesma facilidade com que prevemos o sexto cubo?

Existem três fontes de incerteza, mesmo em mundos de matemática pura: duas fontes óbvias e uma não tão óbvia.

A primeira fonte de incerteza é que mesmo uma criatura de matemática pura, vivendo inserida em um mundo de matemática pura, pode não saber matemática. Os humanos andavam na Terra muito antes de Galileu/Newton/Einstein descobrirem a lei da gravidade que nos impede de sermos lançados no espaço. Embora você possa não estar ciente delas, leis fundamentais estáveis governam a sua existência. Não há nenhuma lei física que exija que essas leis sejam explicitamente conhecidas pelos cérebros submetidos a elas.

Ainda não temos a Teoria de Tudo. Nossas melhores teorias atuais são construções matemáticas, mas não estão perfeitamente integradas entre si. A explicação provável é que, como já se mostrou ser o caso, estamos vendo manifestações superficiais de uma matemática mais profunda. Então, de longe, o melhor palpite é a realidade ser feita de matemática; mas não sabemos totalmente de qual matemática ela é feita, ainda.

Para diferenciar entre as teorias e tornar visível a incerteza remanescente, os físicos precisam construir aceleradores de partículas de grande porte - para manifestar sua incerteza remanescente de alguma forma visível. O fato de os físicos precisarem se esforçar tanto, para não terem certeza, sugere que essa não é a fonte de nossa incerteza sobre os preços no mercado de ações.

A segunda fonte óbvia de incerteza é que, mesmo quando você conhece todas as leis relevantes da física, você pode não ter capacidade computacional suficiente para extrapolá-las. Conhecemos todas as leis físicas fundamentais relevantes para uma cadeia de aminoácidos que se dobra em uma proteína. Mas ainda não conseguimos prever a forma da proteína a partir dos aminoácidos. Uma pequena molécula de 5 nanômetros que se dobra em um microssegundo é informação demais para os computadores atuais lidarem (não importam os tigres nem a bolsa de valores). Nossos esforços pioneiros no dobramento de proteínas empregam técnicas inteligentes, evitando a equação de Schrödinger subjacente. Quando se trata de descrever um objeto de 5 nanômetros usando princípios físicos básicos, como quarks - bem, você nem se dá ao trabalho de tentar.

Para determinar as formas das proteínas, dependemos de instrumentos como cristalografia de raios-x e RMN. Esses métodos nos permitem explorar as estruturas definidas pela física conhecida e por uma sequência de DNA que sabemos. Não somos logicamente oniscientes; não podemos apreender todas as implicações de nossos pensamentos; não sabemos no que acreditamos.

A terceira fonte de incerteza é a mais difícil de entender, e Nick Bostrom escreveu um livro sobre ela. Suponha que a sequência {1, 8, 27, 64, 125… } exista; suponha que isso seja um fato. E suponha que no topo de cada cubo haja uma pessoa pequena - uma pessoa por cubo - e suponha, também, que isso também seja um fato.

Se você estiver do lado de fora e adotar uma perspectiva global - olhando de cima para baixo para a sequência de cubos e as pessoas pequenas empoleiradas no topo - esses dois fatos mostram tudo o que há para saber sobre a sequência e as pessoas.

Contudo, se você é uma das pessoas pequenas empoleiradas no topo de um cubo e conhece esses dois fatos, ainda há uma terceira informação necessária para fazer previsões: “Em qual cubo estou pisando?”

Você espera encontrar-se sobre um cubo, você não espera encontrar-se sobre o número 7. Suas expectativas são limitadas pelo seu conhecimento de física básica; suas crenças são falseáveis. No entanto, ainda é necessário verificar se você está sobre 1.728 ou 5.177.717. Se você puder fazer aritmética mental rapidamente, ver que os dois primeiros dígitos de um cubo de quatro dígitos são 17 será suficiente para deduzir que os últimos dígitos são 2 e 8. Caso contrário, talvez seja necessário descobrir o 2 e o 8 também.

Para compreender a aparência do céu noturno, não basta conhecer as leis da física. Nem mesmo é suficiente ter onisciência lógica sobre suas consequências. Você precisa saber onde está no universo. É preciso saber que a Terra é o ponto de observação do céu noturno. A informação necessária não se limita à localização da Terra no universo visível, mas engloba todo o universo, incluindo as partes invisíveis para nossos telescópios por estarem muito distantes, bem como diferentes universos inflacionários e ramificações alternativas de Everett.

É uma boa aposta que a “incerteza sobre as condições iniciais na fronteira” seja realmente uma in-

certeza indexical. Mas, se não for, trata-se de incerteza empírica, a incerteza sobre o universo é a partir de uma perspectiva global, o que a coloca na mesma classe da incerteza sobre as leis fundamentais. Sempre que nosso melhor palpite é que o "mundo real" tem um componente irremediavelmente confuso, é devido à segunda e à terceira fontes de incerteza - incerteza lógica e incerteza indexical.

A ignorância das leis fundamentais não diz que um padrão de aparência confusa é realmente confuso. Pode ser que você ainda não tenha descoberto a ordem.

Contudo, redes confusas de expressão gênica já encontraram a beleza oculta - o nível estável da física subjacente. Como já encontramos a ordem mestra, podemos supor que não encontraremos nenhum padrão secreto adicional que torne a biologia tão fácil quanto uma sequência de cubos. Conhecendo as regras do jogo, sabemos que o jogo é difícil. Não temos poder de computação suficiente para fazer química de proteínas a partir da física (a segunda fonte de incerteza) e os caminhos evolutivos podem ter seguido caminhos diferentes em planetas diferentes (a terceira fonte de incerteza). Novas descobertas na física básica não nos ajudarão aqui.

Se você fosse um grego na Grécia antiga, analisando os dados brutos de um experimento biológico, buscaria encontrar uma estrutura oculta com uma elegância pitagórica, como todas as proteínas alinhadas em um icosaedro perfeito. Mas, na biologia moderna, sabemos que a elegância pitagórica não está presente na superfície, mas sim em níveis mais profundos que não nos ajudam a superar incertezas indexicais e lógicas.

Da mesma forma, na mecânica quântica, podemos ter certeza de que ninguém pode prever os resultados de certos experimentos, pois a teoria fundamental nos informa que diferentes versões de nós mesmos verão resultados distintos. Se você possui conhecimento das leis fundamentais, sabe que existe uma sequência de cubos, cada um com uma pessoa pequena no topo, que essas pessoas são idênticas, exceto pelo fato de estarem em cubos diferentes, e que você é uma dessas pessoas pequenas, então você sabe que não tem como saber sobre qual cubo está sem olhar para ele.

Nossa compreensão atual sugere que o "mundo real" é uma construção matemática altamente organizada, determinista e vasta, cuja simulação é muito cara. Em vez de prever o próximo cubo numa sequência de cubos, a "realidade" é mais como um grupo de indivíduos dispostos sobre cubos, sem saber quem somos ou ter habilidades mentais aritméticas excepcionais. Embora tenhamos alguma compreensão das regras, elas não nos permitem prever perfeitamente o futuro.

Pronto, isso não parece com a realidade?

Mas a incerteza existe no mapa, não no território. Se ignoramos um fenômeno, isso é um fato sobre nosso estado mental, não um fato sobre o fenômeno em si. Incerteza empírica, incerteza lógica e incerteza indexical são apenas nomes para nossa própria perplexidade. O melhor palpite atual é que o mundo é matemático e a matemática é perfeitamente regular. A bagunça está apenas nos olhos de quem a vê.

Até mesmo o enorme pântano da blogosfera está embutido nessa física perfeita, sendo tão ordenada quanto {1, 8, 27, 64, 125… }.

Portanto, a Internet não é uma grande confusão… é uma série de cubos.

184 — Bela probabilidade

Deveríamos esperar que a racionalidade fosse, em algum nível, simples? Deveríamos procurar e esperar pela beleza subjacente nas artes da crença e da escolha?

Permitam-me introduzir esta questão tomando emprestada uma reclamação do falecido grande Mestre Bayesiano, E. T. Jaynes [1]:

> Dois pesquisadores médicos usam o mesmo tratamento de forma independente, em hospitais diferentes. Nenhum dos dois se rebaixaria a falsificar os dados, mas um havia decidido de antemão que, devido aos recursos finitos, ele pararia após tratar 100 pacientes (n=100), independentemente do número de curas observadas até então. O outro havia apostado sua reputação na eficácia do tratamento, e decidiu que não pararia até ter dados que indicassem uma taxa de cura definitivamente superior a 60%, independentemente do número de pacientes que isso pudesse exigir. Mas, na verdade, ambos pararam exatamente com os mesmos dados: n = 100 [pacientes], r = 70 [curas]. Devemos então tirar conclusões diferentes de seus experimentos?" [Presumivelmente, os dois grupos de controle também tiveram resultados iguais[4].]

Cyan nos direciona para o capítulo 37 do excelente livro de estatística de MacKay, gratuito online, para uma explicação mais completa deste problema. [2]

De acordo com o procedimento estatístico tradicional - que, acredito, ainda é ensinado atualmente - os dois pesquisadores realizaram experimentos diferentes com critérios de parada diferentes. Os dois experimentos poderiam ter terminado com dados diferentes e, portanto, representam testes de hipóteses distintos, que exigem análises estatísticas diferentes. É bem possível que o primeiro experimento seja "estatisticamente significativo", enquanto o segundo não.

O fato de você ficar ou não, incomodado com isso diz muito sobre sua atitude em relação à teoria da probabilidade e, na verdade, à própria racionalidade.

Os estatísticos não bayesianos podem dar de ombros, dizendo: "Bem, nem todas as ferramentas estatísticas têm os mesmos pontos fortes e fracos, sabe - um martelo não é como uma chave de fenda - e se você aplicar ferramentas estatísticas diferentes, poderá obter resultados diferentes, assim como usar os mesmos dados para calcular uma regressão linear ou treinar uma rede neural regularizada. É preciso usar a ferramenta certa para a ocasião. A vida é uma bagunça".

E há também a resposta bayesiana: "Desculpe? O impacto evidencial de um método experimental fixo, que produz os mesmos dados, depende dos pensamentos privados do pesquisador? E você tem a coragem de nos acusar de sermos "muito subjetivos?"

Se a Natureza for de determinada forma, a probabilidade de os dados serem obtidos da forma como

4 NT. Texto original em inglês. *Two medical researchers use the same treatment independently, in different hospitals. Neither would stoop to falsifying the data, but one had decided beforehand that because of finite resources he would stop after treating n = 100 patients, however many cures were observed by then. The other had staked his reputation on the efficacy of the treatment, and decided he would not stop until he had data indicating a rate of cures definitely greater than 60%, however many patients that might require. But in fact, both stopped with exactly the same data: n = 100 [patients], r = 70 [cures]. Should we then draw different conclusions from their experiments?" [Presumably the two control groups also had equal results.]*

vimos será uma coisa. Se a Natureza for de outra forma, a probabilidade de os dados serem obtidos dessa forma será outra. Porém, a probabilidade de um determinado estado da Natureza produzir os dados observados não está relacionada às intenções privadas do pesquisador. Portanto, quaisquer que sejam nossas hipóteses sobre a Natureza, a razão de verossimilhança, o impacto evidencial e a crença posterior devem ser os mesmos entre dois experimentos. Se os dois métodos do Estilo Antigo chegarem a respostas diferentes, pelo menos um deles deve descartar informações relevantes - ou simplesmente fazer o cálculo errado.

A guerra antiga entre bayesianos e frequentistas malditos se estende por décadas, e não vou tentar recontar essa história antiga neste ensaio.

Mas um dos conflitos centrais é que os bayesianos esperam que a teoria da probabilidade seja… qual é a palavra que estou procurando? "Organizada?" "Limpa?" "Autoconsistente?"

Como diz Jaynes, os teoremas da probabilidade bayesiana são apenas isso, teoremas em um sistema de prova coerente. Não importa quais derivações você use, em qual ordem, os resultados da teoria da probabilidade Bayesiana sempre devem ser consistentes – cada teorema compatível com todos os outros teoremas.

Se quiser saber a soma de 10 + 10, você pode redefini-la como (2 × 5) + (7 + 3) ou como (2 x (4 + 6) ou usar quaisquer outros truques legais de sua preferência, mas o resultado sempre tem que sair o mesmo, neste caso, 20. Se der 20 de uma maneira e 19 da outra, então você pode concluir que fez algo ilegal em pelo menos uma das duas ocasiões. (Na aritmética, a operação ilegal é geralmente a divisão por zero; na teoria da probabilidade, geralmente é um infinito que não foi considerado o limite de um processo finito).

Se você obtiver o resultado 19 = 20, procure com atenção o erro que acabou de cometer, porque é improvável que você tenha destruído a própria aritmética. Se alguém conseguir derivar uma contradição real da teoria da probabilidade Bayesiana – como, digamos, dois impactos evidenciais diferentes do mesmo método experimental produzindo os mesmos resultados – então todo o edifício se transforma em fumaça. Junto com a teoria dos conjuntos, porque tenho certeza que a ZF[5] fornece um modelo para a teoria da probabilidade.

Matemática! Essa é a palavra que eu estava procurando. Bayesianos esperam que a teoria da probabilidade seja matemática. É por isso que estamos interessados no Teorema de Cox[6] e em suas muitas extensões, mostrando que qualquer representação de incerteza que obedeça a certas restrições deve ser mapeada na teoria da probabilidade. A matemática coerente é ótima, mas a matemática única é ainda melhor.

E ainda assim… a racionalidade deveria ser matemática? Não é de forma alguma uma conclusão precipitada que a probabilidade deve ser bonita. O mundo real é confuso - então você não deveria precisar um raciocínio confuso para lidar com ele? Talvez os estatísticos não bayesianos, com sua vasta coleção de métodos e justificativas ad-hoc, sejam estritamente mais competentes porque possuem uma caixa de ferramentas estritamente maior. É bom quando os problemas são claros, mas geralmente não são, e você tem que viver com isso.

Afinal de contas, é um fato bem conhecido que você não pode usar métodos Bayesianos em muitos problemas porque o cálculo Bayesiano é computacionalmente intratável. Então, por que não deixar muitas flores desabrocharem? Por que não ter mais de uma ferramenta em sua caixa de ferramentas?

Essa é a diferença fundamental na mentalidade. Os estatísticos da velha guarda pensavam em termos

5 NT. A Teoria de Zermelo-Fraenkel (ZF) é a formulação axiomática mais influente da teoria dos conjuntos, servindo de base para grande parte da matemática moderna. Desenvolvida no início do século XX por Ernst Zermelo e Abraham Fraenkel, estabelece axiomas para a formação e manipulação de conjuntos, evitando paradoxos como o de Russell.

6 NT. O Teorema de Cox, proposto pelo físico Richard Threlkeld Cox, mostra que qualquer sistema de raciocínio que represente incerteza de modo lógico e consistente se equivale à teoria de probabilidades. Ou seja, a probabilidade surge como a única forma coerente de quantificar crenças, tornando-se fundamental para a inferência racional em situações incertas.

de ferramentas, truques para lidar com problemas específicos. Os bayesianos - pelo menos este bayesiano, embora eu não ache que estou falando apenas por mim - pensamos em termos de leis.

Procurar leis não é o mesmo que procurar ferramentas especialmente elegantes e bonitas. A Segunda Lei da Termodinâmica[7] não é uma geladeira especialmente arrumada e bonita.

O ciclo de Carnot é um motor ideal - na verdade, o motor ideal. Nenhum motor alimentado por dois reservatórios de calor pode ser mais eficiente do que um motor de Carnot[8]. Como corolário, todos os motores termodinamicamente reversíveis operando entre os mesmos reservatórios de calor são igualmente eficientes.

Mas, é claro, você não pode usar um motor de Carnot para alimentar um carro de verdade. O motor de um carro real tem a mesma semelhança com um motor de Carnot que os pneus do carro têm com cilindros rolantes perfeitos.

Claramente, então, um motor de Carnot é uma ferramenta inútil para construir um carro real. A Segunda Lei da Termodinâmica, obviamente, não se aplica aqui. É muito difícil fazer um motor que a obedeça, no mundo real. Esse é o tipo de confusão que, na minha opinião, reina sobre aqueles que ainda se apegam aos Velhos Costumes.

Não, você nem sempre pode fazer o cálculo bayesiano exato para um problema. Às vezes você deve buscar uma aproximação; muitas vezes, de fato. Isso não significa que a teoria da probabilidade deixou de se aplicar, assim como sua incapacidade de calcular a aerodinâmica de um Boeing 747, átomo por átomo, não implica que o 747 não é feito de átomos. Independentemente da aproximação que você use, ela funciona enquanto se aproxima do cálculo bayesiano ideal - e falha enquanto se afasta.

As provas de coerência e unicidade do Bayesianismo são bidirecionais. Da mesma forma que qualquer cálculo que obedeça aos axiomas de coerência de Cox (ou a qualquer uma das muitas reformulações e generalizações) deve ser mapeado em probabilidades, qualquer abordagem não bayesiana falhará em um dos testes de coerência. Isso pode levar a punições como *Dutch-booking* (aceitar combinações de apostas que são perdas certas ou rejeitar combinações de apostas que são ganhos certos).

Você pode não conseguir calcular a resposta ideal. Mas independentemente da aproximação que você usar, tanto seus fracassos quanto seus sucessos serão explicáveis em termos da teoria da probabilidade bayesiana. Você pode não saber a explicação; isso não significa que não exista explicação.

Então, você quer usar uma regressão linear, em vez de fazer atualizações bayesianas? Mas observe a estrutura subjacente da regressão linear e você verá que ela corresponde a escolher a melhor estimativa pontual dada uma função de verossimilhança Gaussiana e um prior uniforme sobre os parâmetros.

Você quer usar uma regressão linear regularizada, porque isso funciona melhor, na prática? Bem, isso corresponde (diz o bayesiano) a ter um a priori gaussiano sobre os pesos.

Às vezes você não pode usar os métodos Bayesianos literalmente; muitas vezes, de fato. Mas quando você pode usar o cálculo bayesiano exato que usa cada fragmento de conhecimento disponível, está feito. Você nunca encontrará um método estatístico que produza uma resposta melhor. Você pode encontrar uma aproximação barata que funcione excelentemente quase o tempo todo, e será mais barata, mas não será mais precisa. Não, a menos que o outro método use conhecimento, talvez na forma de informações anteriores disfarçadas, que você não está permitindo no cálculo bayesiano; e então, quando você inserir as informa-

---

7 NT. **Segunda Lei da Termodinâmica**: Princípio que afirma que a entropia (medida de desordem) de um sistema isolado tende a aumentar com o tempo, indicando a irreversibilidade de processos naturais. Estabelece que a energia se dispersa espontaneamente e sistemas evoluem para estados de maior equilíbrio térmico, impossibilitando a conversão integral de calor em trabalho sem perdas. Define a direcionalidade intrínseca dos fenômenos termodinâmicos.

8 NT. **Motor de Carnot**: Máquina térmica teórica idealizada por Sadi Carnot, que opera em um ciclo reversível entre duas fontes de temperatura (quente e fria). Considerada o modelo de máxima eficiência termodinâmica, sua performance depende exclusivamente da diferença entre as temperaturas das fontes. Serve como referência para avaliar limites de conversão de calor em trabalho em sistemas reais, conforme a Segunda Lei da Termodinâmica.

ções anteriores no cálculo bayesiano, o cálculo bayesiano será novamente igual ou superior.

Ao utilizar uma ferramenta estatística ad-hoc tradicional com uma justificativa ad-hoc (embora frequentemente bastante interessante), é impossível prever se alguém apresentará uma ferramenta ainda mais inteligente no futuro. Contudo, quando é possível usar diretamente um cálculo que reflete a lei bayesiana, a situação muda drasticamente, semelhante a instalar um motor térmico de Carnot em um carro. Como diz o ditado, "Bayes-ótimo".

Parece que os toolboxers estão olhando para a sequência de cubos {1, 8, 27, 64, 125…} e apontando as primeiras diferenças {7, 19, 37, 61…}, e dizendo: "Veja, a vida não é sempre organizada - você precisa se adptar às circunstâncias." E os Bayesianos estão apontando para as terceiras diferenças, o nível estável subjacente {6, 6, 6, 6, 6…}. E os críticos estão dizendo: "De que diabos você está falando? São 7, 19, 37, não 6, 6, 6. Você está simplificando demais esse problema complexo; você está muito apegado à simplicidade."

Não é necessariamente simples em um nível superficial. É preciso mergulhar mais profundamente para encontrar a estabilidade.

Pense em leis, não em ferramentas. A necessidade de calcular aproximações para uma lei não altera a própria lei. Os aviões ainda são compostos por átomos, eles não são governados por exceções especiais na natureza para cálculos aerodinâmicos. A aproximação existe no mapa, não no território. Você pode conhecer a Segunda Lei da Termodinâmica e, ainda assim, se candidatar a uma vaga como engenheiro para construir um motor de carro imperfeito. A Segunda Lei não deixa de ser aplicável; seu conhecimento dessa lei e dos ciclos de Carnot o ajuda a se aproximar o máximo possível da eficiência ideal.

Não somos cativados pelos métodos Bayesianos apenas por sua beleza. A beleza é um efeito colateral. Os teoremas Bayesianos são elegantes, coerentes, ótimos e comprovadamente únicos porque são leis.

Referências

[1] Edwin T. Jaynes, "Probability Theory as Logic," in Maximum Entropy and Bayesian Methods, ed. Paul F. Fougère (Springer Netherlands, 1990).

[2] David J. C. MacKay, Information Theory, Inference, and Learning Algorithms (New York: Cambridge University Press, (2003).

185 — Fora do laboratório

"Fora do laboratório, os cientistas não são mais sábios do que ninguém." Às vezes, este provérbio é dito por cientistas, com humildade e tristeza, para se lembrarem da sua própria falibilidade. Às vezes, esse provérbio é dito por razões bem menos louváveis, para desvalorizar conselhos indesejados de especialistas. O provérbio é verdadeiro? Provavelmente não em um sentido absoluto. Parece demasiado pessimista dizer que os cientistas não são literalmente mais sábios do que a média, que a correlação é literalmente zero.

Mas o provérbio parece verdadeiro até certo ponto, e proponho que fiquemos muito perturbados com esse fato. Não devemos suspirar e balançar a cabeça tristemente. Em vez disso, devemos nos sentar eretos em alarme. Por quê? Bem, suponha que um pastor aprendiz seja meticulosamente treinado para contar ovelhas enquanto entram e saem de um curral. Por meio desse treinamento, o pastor sabe quando todas as ovelhas partiram e retornaram. Então você oferece maçãs ao pastor e pergunta "Quantas maçãs?", mas o pastor fica olhando para você sem entender nada, porque ele não foi treinado para contar maçãs - apenas ovelhas. Essa situação pode levar à suspeita de que o pastor não compreenda completamente o conceito de contagem.

Agora, imagine que descobrimos que um economista com doutorado compra um bilhete de loteria toda semana. Temos que nos perguntar: será que esse indivíduo compreende realmente o conceito de utilidade esperada, em um nível instintivo? Ou ele apenas foi treinado para executar certos truques de álgebra?

Podemos pensar na história de Richard Feynman sobre um programa de ensino de física que fracassou:

> Os alunos haviam memorizado tudo, mas não sabiam o significado de nada. Quando ouviam "luz que é refletida de um meio com um índice", não sabiam que isso se referia a um material como a água. Não sabiam que a "direção da luz" significava a direção em que você vê algo quando olha para ele, e assim por diante. Tudo estava na memória deles, mas nada era traduzido em palavras significativas. Por exemplo, se eu perguntasse: "O que é o ângulo de Brewster?", estou entrando com as palavras-chave corretas no computador. Mas se eu dissesse: "Olhe para a água", nada acontecia, pois eles não tinham nada em sua memória sobre "Olhe para a água"[9].

Suponha que temos um cientista aparentemente competente, capaz de projetar um experimento com N sujeitos; os N indivíduos recebem um tratamento randomizado; juízes cegos avaliam os resultados dos sujeitos; e depois os resultados são processados em um computador para verificar se são estatisticamente significativos a um nível de confiança de 0,05. Essa não é uma mera tradição ritualizada. Isso não é um capricho arbitrário, como escolher o garfo certo para a salada. Isso é uma tradição ritualizada para testar hipóteses experimentalmente. Por que devemos testar nossas hipóteses experimentalmente? É porque sabemos que as revistas científicas exigem isso antes de publicar nossos artigos? Porque fomos ensinados a fazer isso na faculdade? Porque todos repetem em uníssono que é importante realizar o experimento e que vão olhar para você de maneira estranha se disser o contrário?

---

9 NT. Texto original em inglês. The students had memorized everything, but they didn't know what anything meant. When they heard "light that is reflected from a medium with an index," they didn't know that it meant a material such as water. They didn't know that the "direction of the light" is the direction in which you see something when you're looking at it, and so on. Everything was entirely memorized, yet nothing had been translated into meaningful words. So if I asked, "What is Brewster's Angle?" I'm going into the computer with the right keywords. But if I say, "Look at the water," nothing happens—they don't have anything under "Look at the water"!

Não: porque, para mapear um território, você deve sair e observar o território. Não é possível criar um mapa preciso de uma cidade enquanto está sentado na sala de estar de olhos fechados, tendo pensamentos agradáveis sobre como você gostaria que a cidade fosse. É preciso sair, percorrer a cidade e registrar suas observações no papel para representar fielmente o que vê. Isso acontece em uma escala menor toda vez que você verifica se seus cadarços estão desamarrados. Os fótons chegam do sol, refletem em seus cadarços, atingem sua retina, são transduzidos em frequências de disparo neural e são reconstruídos pelo córtex visual em um padrão de ativação altamente correlacionado com o formato atual de seus cadarços. Para adquirir novas informações sobre o território, você deve interagir com ele. Deve haver algum processo físico real que correlacione o estado de seu cérebro com o estado do ambiente. Os processos de raciocínio não são mágicos; você pode fornecer descrições causais de como eles funcionam. Isso tudo quer dizer que, para descobrir as coisas, é necessário observar.

Agora, o que devemos pensar de um cientista que parece competente no laboratório, mas que, fora dele, acredita em um mundo espiritual? Devemos nos perguntar por quê, e o cientista pode responder algo como: “Bem, ninguém realmente sabe, e eu admito que não tenho nenhuma evidência - é uma crença religiosa, não pode ser refutada garantidamente por meio da observação.” Não posso evitar concluir que essa pessoa literalmente não compreende realmente a razão pela qual é necessário observar as coisas. Ela pode ter aprendido um ritual específico de experimentação, mas não entende a razão disso - que para mapear um território, é necessário observá-lo; que para obter informações sobre o ambiente, você precisa se envolver em um processo causal por meio do qual você interage com o ambiente e acaba se correlacionando com ele. Essa lógica se aplica tanto a um projeto experimental duplo-cego que coleta informações sobre a eficácia de um novo dispositivo médico quanto a seus olhos que coletam informações sobre os seus cadarços.

Talvez nosso cientista espiritual diga: “Mas não se trata de experimentação. Os espíritos falam comigo em meu coração.” Bem, se realmente supusermos que os espíritos estão se comunicando de alguma forma, isso é uma interação causal e conta como uma observação. A teoria da probabilidade ainda se aplica. Se você propõe que uma experiência pessoal de “vozes dos espíritos” é uma evidência de espíritos reais, você precisa propor que haja uma razão de verossimilhança favorável para os espíritos causarem “vozes de espíritos”, em comparação com outras explicações para “vozes de espíritos”, o que é suficiente para superar a improbabilidade inerente a uma crença complexa e multifacetada. Não perceber que “os espíritos falam comigo em meu coração” é uma forma de “interação causal” é análogo a um estudante de física que não percebe que um “meio com um índice” se refere a um material como a água.

É fácil ser enganado, talvez, pelo fato de que as pessoas usando jalecos de laboratório mencionam “interação causal”, enquanto aquelas com joias extravagantes falam sobre “espíritos falando”. Os debatedores que usam roupas diferentes, como todos sabemos, demarcam esferas independentes de existência - “magistérios separados”, segundo a frase imortal de Stephen J. Gould. Na verdade, “interação causal” é apenas uma maneira elegante de dizer “algo que faz outra coisa acontecer”, e a teoria da probabilidade não se importa com as roupas que você veste.

Na sociedade moderna, existe uma ideia predominante de que questões espirituais não podem ser resolvidas por lógica ou observação, e que as pessoas podem ter as crenças religiosas que desejarem. Se um cientista cai nessa armadilha e decide viver sua vida extralaboratorial de acordo com isso, então, para mim, isso sugere que ele apenas entende o princípio experimental como uma convenção social. Ele sabe quando se espera que conduza experimentos e teste resultados em busca de significância estatística. Mas, quando está em um contexto em que é socialmente aceitável inventar crenças malucas sem fundamento, ele faz isso com a mesma facilidade.

O pastor aprendiz é informado de que se “sete” ovelhas saíram e “oito” ovelhas saíram, então “quinze” ovelhas devem ter retornado. Por que “quinze” em vez de “quatorze” ou “três”? Porque, do contrário, você não terá jantar hoje, é por isso! Portanto, esse é um tipo de treinamento profissional que funciona de certa forma - mas, se a única razão pela qual “sete ovelhas mais oito ovelhas são quinze ovelhas” é uma convenção social, então talvez “sete maçãs mais oito maçãs são três maçãs”. Quem pode dizer que as regras não devem diferir para as maçãs?

Mas se você compreende por que as regras funcionam, perceberá que a operação matemática de

adição é a mesma para ovelhas e maçãs. Isaac Newton é justamente reverenciado, não tanto por sua teoria ultrapassada da gravidade, mas por descobrir que - de forma espantosa e surpreendente - os planetas celestes, nos céus gloriosos, obedecem exatamente às mesmas regras que as maçãs que caem na Terra. No mundo macroscópico - o ambiente cotidiano ancestral - diferentes árvores produzem diferentes frutos, costumes diferentes se aplicam a pessoas diferentes em momentos diferentes. Um universo verdadeiramente unificado, com leis universais invariáveis, é uma noção altamente contraintuitiva para os seres humanos! Somente os cientistas realmente acreditam nisso, embora algumas religiões possam falar eloquentemente sobre a "unidade de todas as coisas".

Como disse Richard Feynman:

> Quando olhamos um copo de perto, vemos o universo inteiro. Existem os fenômenos físicos: o líquido que se retorce e evapora dependendo do vento e do clima, os reflexos no vidro e nossa imaginação adiciona os átomos. O vidro é uma destilação das rochas da Terra, e em sua composição, encontramos o segredo da idade do universo e a evolução das estrelas. Que coleção estranha de substâncias químicas existe no vinho? Como elas surgiram? Há fermento, enzimas, substratos e produtos. No vinho, encontramos uma grande generalização: toda vida é fermentação. Ninguém pode entender a química do vinho sem compreender, como Louis Pasteur fez, a causa de muitas doenças. Como é vívida a cor do clarete, enquanto se projeta na consciência de quem o observa! Se nossas mentes pequenas, por mera conveniência, dividem essa taça de vinho, esse universo, em partes - física, biologia, geologia, astronomia, psicologia e assim por diante - lembremos que a natureza não faz essa divisão! Portanto, devemos unificar tudo novamente, sem esquecer, em última análise, qual é o propósito. Que ela nos dê mais um prazer final: beber e esquecer tudo![10]

Algumas religiões, especialmente aquelas inventadas ou reformuladas após Isaac Newton, podem professar que "tudo está interconectado". (Como há um isomorfismo óbvio entre gráficos e seus complementos, essa profunda sabedoria transmite exatamente a mesma informação útil que um gráfico sem arestas). Mas quando se trata da essência real da religião, profetas e sacerdotes seguem a prática humana antiga de inventar tudo à medida que avançam. E eles estabelecem uma regra para mulheres com menos de doze anos, outra regra para homens com mais de treze anos; uma regra para o Sabbath e outra para os dias de semana; uma regra para a ciência e outra para a feitiçaria…

A realidade, aprendemos para nosso choque, não consiste em uma coleção de magistérios separados, mas sim em um único processo unificado governado por regras matemáticas simples em níveis mais baixos. Edifícios diferentes em um campus universitário não pertencem a universos diferentes, embora às vezes possa parecer assim. O universo não está dividido entre mente e matéria, ou vida e não vida; os átomos em nossas cabeças interagem perfeitamente com os átomos do ar ao nosso redor. O Teorema de Bayes também não varia de um lugar para outro.

Portanto, se um determinado cientista, fora de sua área de especialização, é tão suscetível a ideias irracionais quanto qualquer outra pessoa, talvez nunca tenha compreendido por que as regras científicas funcionam. Talvez ele pode conseguir recitar algumas noções do falsificacionismo de Popper[11], mas não tem um entendimento profundo, seja no nível algébrico da teoria da probabilidade ou no nível causal da cognição como um processo de máquina. Ele pode ter sido treinado para seguir um certo protocolo no laboratório,

---

10 NT. Texto original em inglês. *If we look at a glass closely enough we see the entire universe. There are the things of physics: the twisting liquid which evaporates depending on the wind and weather, the reflections in the glass, and our imagination adds the atoms. The glass is a distillation of the Earth's rocks, and in its composition we see the secret of the universe's age, and the evolution of the stars. What strange array of chemicals are there in the wine? How did they come to be? There are the ferments, the enzymes, the substrates, and the products. There in wine is found the great generalization: all life is fermentation. Nobody can discover the chemistry of wine with- out discovering, as did Louis Pasteur, the cause of much disease. How vivid is the claret, pressing its existence into the consciousness that watches it! If our small minds, for some convenience, divide this glass of wine, this universe, into parts—physics, biology, geology, astronomy, psychology, and so on—remember that Nature does not know it! So let us put it all back together, not forgetting ultimately what it is for. Let it give us one more final pleasure: drink it and forget it all!*

11 NT. No **falsificacionismo**, proposto por Karl Popper, uma teoria científica deve ser testável e passível de refutação, sendo aceito que, enquanto resistir a tentativas de falseamento, ela permanece válida. Em contrapartida, teorias que não possam ser postas à prova não são consideradas científicas.

mas não valoriza a importância das evidências; quando vai para casa, tira o jaleco de laboratório e relaxam com alguma bobagem confortável. E sim, isso me faz questionar se posso confiar nas opiniões desse cientista mesmo em sua própria área - especialmente quando se trata de qualquer questão controversa, qualquer questão em aberto, qualquer coisa que ainda não foi confirmado por evidências robustas e convenções sociais.

Talvez possamos superar o provérbio - ser racionais em nossas vidas pessoais, não apenas em nossas vidas profissionais. Não devemos permitir que um simples provérbio nos detenha, como disse Voltaire: “Um ditado espirituoso não prova nada”. Talvez possamos alcançar um patamar mais elevado se estudarmos a teoria da probabilidade o suficiente para entender por que as regras funcionam e a psicologia experimental o suficiente para ver como elas se aplicam aos casos do mundo real - se pudermos aprender a observar a água. Uma ambição como essa não tem a modéstia confortável de poder confessar que, fora de nossa especialização, não somos melhores do que ninguém. Mas se nossas teorias de racionalidade não se generalizam para a vida cotidiana, estamos cometendo um erro. Não existem dois universos separados: um dentro e outro fora do laboratório.

186 — A Segunda Lei da Termodinâmica e os motores da cognição

A Primeira Lei da Termodinâmica, mais conhecida como Conservação da Energia, diz que não se pode criar energia do nada: ela proíbe as máquinas de movimento perpétuo do primeiro tipo, que funcionam indefinidamente sem consumir combustível ou qualquer outro recurso. De acordo com nossa visão moderna da física, a energia é conservada em cada interação individual de partículas. Por indução matemática, vemos que não importa o tamanho de um conjunto de partículas, ele não pode produzir energia do nada — não sem violar o que atualmente acreditamos serem as leis da física.

É por isso que o Escritório de Patentes dos EUA rejeitará sumariamente sua proposta incrivelmente engenhosa de um conjunto de rodas e engrenagens que faz com que uma mola contraia outra à medida que a primeira se expande, e assim continue a realizar trabalho indefinidamente, de acordo com seus cálculos. Existe uma prova totalmente geral de que pelo menos uma roda deve violar (nosso modelo padrão) das leis da física para isso acontecer. Então, a menos que você possa explicar como uma roda viola as leis da física, o conjunto de rodas também não pode fazê-lo.

Um argumento semelhante é aplicável a um "impulso sem reação", um sistema de propulsão que viola o Princípio de Conservação do Momento. Na física padrão, o momento é conservado para todas as partículas individuais e suas interações. Por indução matemática, o momento é conservado para sistemas físicos, independentemente de seu tamanho. Se você puder visualizar duas partículas colidindo uma contra a outra e sempre saindo com o mesmo momento total com que começaram, então poderá ver como o aumento da escala de partículas para uma coleção gigantesca e complexa de engrenagens não mudará nada. Mesmo que haja um trilhão de quatrilhões de átomos envolvidos, 0 + 0 +. . . + 0 = 0.

Mas a Conservação de Energia, como tal, não pode proibir a conversão de calor em trabalho. Na verdade, é possível construir uma caixa selada que converta cubos de gelo e eletricidade armazenada em água quente. O processo não é difícil. A energia não pode ser criada nem destruída, então a mudança líquida de energia, da transformação (cubos de gelo + eletricidade) para (água quente) deve ser 0. Portanto, não poderia violar a Conservação de Energia se o processo fosse feito ao contrário...

As máquinas de movimento perpétuo do segundo tipo, que convertem água quente em corrente elétrica e cubos de gelo, são proibidas pela Segunda Lei da Termodinâmica. A segunda lei é um pouco mais difícil de entender, por ser essencialmente bayesiana por natureza.

Sim, é isso mesmo.

A lei física essencial subjacente à Segunda Lei da Termodinâmica é um teorema que pode ser comprovado no modelo padrão da física: no desenvolvimento ao longo do tempo de qualquer sistema fechado, o volume do espaço de fase[12] é conservado. Imagine segurar uma bola bem acima do chão. Esse cenário pode ser descrito como um ponto em um espaço multidimensional, onde pelo menos uma das dimensões é a "altura da bola acima do solo". Então, quando você solta a bola, ela se move, assim como o ponto multidimensional no espaço de fase que representa todo o sistema, incluindo você e a bola. Na linguagem da física, "espaço de fase" significa que existem dimensões não apenas para a posição das partículas, mas também para seu

12 NT. **Espaço de Fase**: Espaço multidimensional em que todas as posições e momentos possíveis de um sistema são representados.

momento. Por exemplo, um sistema de duas partículas teria 12 dimensões: 3 dimensões para a posição de cada partícula e 3 dimensões para o momento de cada partícula.

Se você tivesse um espaço multidimensional, onde cada dimensão descrevesse a posição de uma engrenagem em um enorme conjunto de engrenagens, então, à medida que você girasse as engrenagens, um único ponto mergulharia e seria arremessado em um espaço de fase de dimensão bastante elevada. Ou seja, assim como você pode considerar uma máquina grande e complexa como um único ponto em um espaço com muitas dimensões, também é possível visualizar as leis da física que descrevem o comportamento dessa máquina ao longo do tempo como uma descrição da trajetória desse ponto no espaço de fase.

A Segunda Lei da Termodinâmica é consequência de um teorema que pode ser provado no modelo padrão da física: se você pegar um volume do espaço de fase e o desenvolver ao longo do tempo usando a física padrão, o volume total do espaço de fase é conservado.

Por exemplo, sejam dois sistemas, X e Y, onde X tem 8 estados possíveis, Y tem 4 estados possíveis, e o sistema conjunto (X; Y) tem 32 estados possíveis.

O desenvolvimento do sistema conjunto ao longo do tempo pode ser descrito como uma regra que mapeia pontos iniciais em pontos futuros. Por exemplo, o sistema poderia começar em X_7Y_2 e, em seguida, desenvolver-se (sob um conjunto de leis físicas) para o estado X_3Y_3 um minuto depois. Isso quer dizer: se X começou no estado X_7 e Y começou no estado Y_2, e observássemos por 1 minuto, veríamos X ir para X_3 e Y ir para Y_3. Essas são as leis da física.

Em seguida, delimitemos um subespaço S do estado do sistema conjunto. O espaço S será o subespaço limitado por X estando no estado X_1 e Y estando nos estados Y_1 a Y_4. Portanto, o volume total de S é de 4 estados.

E suponhamos que, sob as leis da física que governam (X, Y), os estados inicialmente em S se comportam da seguinte maneira:

$$X_1Y_1 \rightarrow X_2Y_1$$
$$X_1Y_2 \rightarrow X_4Y_1$$
$$X_1Y_3 \rightarrow X_6Y_1$$
$$X_1Y_4 \rightarrow X_8Y_1$$

Isso, em poucas palavras, é como funciona um refrigerador.

O subsistema X começou em uma região estreita do espaço de estados - o único estado X_1, na verdade - e Y começou distribuído sobre uma região mais ampla do espaço, estados Y_1 a Y_4. Ao interagirem entre si, Y foi para uma região estreita, e X acabou em uma região ampla; mas o volume total do espaço de fase foi conservado. Quatro estados iniciais foram mapeados para quatro estados finais.

Claramente, desde que o volume total do espaço de fase seja conservado pela física ao longo do tempo, você não pode comprimir Y com mais força do que X se expande, ou vice-versa - pois cada subsistema que sofre uma compressão e ocupa uma região mais estreita do espaço de estados, algum outro subsistema deve se expandir e ocupar uma região mais ampla do mesmo espaço. Agora, digamos que estamos incertos sobre o sistema conjunto (X,Y), e que nossa incerteza é descrita por uma distribuição equiprovável sobre S. Ou seja, estamos bastante certos de que X está no estado X_1, mas Y tem igual probabilidade de estar em qualquer um dos estados Y_1 a Y_4. Se fecharmos os olhos por um minuto e depois os abrirmos novamente, esperaremos ver Y no estado Y_1, mas X pode estar em qualquer um dos estados X_2 a X_8. Na verdade, X só pode estar em alguns dos estados de X_2 a X_8, mas seria muito custoso pensar exatamente quais estados esses poderiam ser, então

diremos simplesmente de X_2 a X_8.

Se considerarmos a entropia de Shannon da nossa incerteza sobre X e Y como sistemas individuais, X começou com 0 bits de entropia porque tinha um único estado definido, e Y começou com 2 bits de entropia porque tinha igual probabilidade de estar em qualquer um dos 4 estados possíveis. (Não há informação mútua entre X e Y.) Um pouco de física ocorreu e, veja só, a entropia de Y foi para 0, mas a entropia de X foi para bits. Assim, a entropia foi transferida de um sistema para outro e diminuiu no subsistema Y; mas devido ao custo de contabilização, não nos preocupamos em rastrear algumas informações e, consequentemente (da nossa perspectiva), a entropia geral aumentou.

Suponha que houvesse um processo físico que mapeasse estados passados em estados futuros dessa forma:

$$X_2Y_1 \rightarrow X_2Y_1$$
$$X_2Y_2 \rightarrow X_2Y_1$$
$$X_2Y_3 \rightarrow X_2Y_1$$
$$X_2Y_4 \rightarrow X_2Y_1$$

Então você poderia ter um processo físico que realmente diminuiria a entropia, porque independentemente de onde você começasse, você terminaria no mesmo lugar. As leis da física, desenvolvendo-se ao longo do tempo, comprimiriam o espaço de fase.

Mas existe um teorema, o Teorema de Liouville, que pode ser provado verdadeiro em relação às nossas leis da física, que diz que isso nunca acontece: o espaço de fase é conservado.

A Segunda Lei da Termodinâmica é um corolário do Teorema de Liouville: *Não importa quão inteligente seja sua configuração de rodas e engrenagens, você nunca conseguirá diminuir a entropia em um subsistema sem a aumentar em outro lugar. Quando o espaço de fase de um subsistema diminui, o espaço de fase de outro subsistema deve aumentar e o espaço conjunto mantém o mesmo volume.*

Acontece que o que era inicialmente um espaço de fase compacto, pode desenvolver curvas, ondulações e convoluções, de modo que, para traçar uma fronteira simples ao redor de toda essa bagunça, você precisa desenhar uma fronteira muito maior do que antes - é isso que dá a aparência de aumento da entropia. (E nos sistemas quânticos, onde diferentes universos seguem caminhos distintos, a entropia realmente aumenta em qualquer universo local. Mas deixaremos essa complicação de lado por enquanto.)

A Segunda Lei da Termodinâmica é, na verdade, de natureza probabilística - se você perguntar sobre a probabilidade de a água quente entrar espontaneamente no estado “água fria e eletricidade”, a probabilidade existe, mas é muito pequena. Isso não significa que o Teorema de Liouville seja violado com probabilidade pequena; afinal, um teorema é um teorema. Isso significa que, se você estiver em um grande volume de espaço de fase no início, mas não souber onde, poderá avaliar uma pequena probabilidade de acabar em algum volume de espaço de fase específico. Até onde você sabe, com probabilidade infinitesimal, esse copo específico de água quente pode ser do tipo que se transforma espontaneamente em corrente elétrica e cubos de gelo. (Desconsiderando, como de costume, os efeitos quânticos).

Portanto, a Segunda Lei é realmente inerentemente bayesiana. Ao se tratar de qualquer sistema termodinâmico real, é uma declaração estritamente legal das suas crenças sobre o sistema, mas apenas uma declaração probabilística sobre o próprio sistema.

“Espere aí”, você diz. “Não foi isso que aprendi nas aulas de física”, você argumenta. “Nas aulas que assisti, a termodinâmica trata de, você sabe, temperaturas. A incerteza é um estado mental subjetivo! A temperatura de um copo d’água é uma propriedade objetiva da água! O que o calor tem a ver com probabilidade?”

Ah, você de pouca confiança.

Em uma direção, a conexão entre calor e probabilidade é relativamente direta: se o único fato que você conhece sobre um copo d'água é sua temperatura, então você está muito mais incerto sobre um copo de água quente do que um copo de água fria.

O calor é o movimento agitado de muitas moléculas minúsculas; quanto mais quentes elas estão, mais rápido podem se mover. Nem todas as moléculas na água quente viajam na mesma velocidade - a "temperatura" não é uma velocidade uniforme de todas as moléculas, mas uma velocidade média das moléculas, que por sua vez corresponde a uma distribuição estatística previsível de velocidades - de qualquer forma, o ponto é que, quanto mais quente a água, mais rápido as moléculas de água podem estar se movendo e, portanto, mais incerto você está sobre a velocidade (não apenas a rapidez) de qualquer molécula individual. Quando você multiplica suas incertezas sobre todas as moléculas individuais, você ficará exponencialmente mais incerto sobre o copo d'água inteiro.

Pegamos o logaritmo desse volume exponencial de incerteza e chamamos isso de entropia. Então, está tudo explicado, entende?

A conexão na outra direção é menos óbvia. Suponha que houvesse um copo d'água sobre o qual, inicialmente, você só soubesse que sua temperatura era de 22 graus. De repente, São Laplace revela a você as localizações e velocidades exatas de todos os átomos na água. Agora você conhece perfeitamente o estado da água, então, pela definição de entropia da teoria da informação, sua entropia é zero. Isso torna sua entropia termodinâmica zero? A água ficou mais fria porque sabemos mais sobre ela?

Ignorando o aspecto quântico por um momento, a resposta é: Sim! Sim, ficou!

Maxwell certa vez perguntou: Por que não podemos pegar um gás uniformemente quente, dividi-lo em dois volumes A e B, e deixar apenas moléculas rápidas passarem de B para A, enquanto apenas moléculas lentas são permitidas passar de A para B? Se você pudesse construir uma porta assim, logo teria gás quente no lado A e gás frio no lado B. Isso seria uma maneira barata de refrigerar alimentos, certo?

O agente que inspeciona cada molécula de gás e decide se deve deixá-la passar é conhecido como "Demônio de Maxwell[13]". E a razão pela qual você não pode construir um refrigerador eficiente dessa maneira é que o Demônio de Maxwell gera entropia no processo de inspecionar as moléculas de gás e decidir quais deixar passar.

Mas e se você já soubesse onde todas as moléculas de gás estavam?

Então você realmente poderia operar o Demônio de Maxwell e extrair trabalho útil.

Assim (novamente ignorando efeitos quânticos por enquanto), se você conhece os estados de todas as moléculas em um copo de água quente, ele está frio em um sentido genuinamente termodinâmico: você pode extrair eletricidade dele e deixar para trás um cubo de gelo.

Isso não viola o Teorema de Liouville, porque se Y é a água, e você é o Demônio de Maxwell (denotado M), o processo físico se comporta assim:

$$M_1 Y_1 \rightarrow M_1 Y_1$$
$$M_2 Y_2 \rightarrow M_2 Y_1$$
$$M_3 Y_3 \rightarrow M_3 Y_1$$
$$M_4 Y_4 \rightarrow M_4 Y_1$$

13 NT. **Demônio de Maxwell**: Experimento mental proposto por James Clerk Maxwell, que descreve um hipotético ser capaz de separar moléculas rápidas e lentas, aparentemente violando a Segunda Lei da Termodinâmica.

Como o Demônio de Maxwell conhece o estado exato de Y, essa é a informação mútua entre M e Y. A informação mútua diminui a entropia conjunta de (M,Y): temos H(M,Y) = H(M) + H(Y) - I(M,Y). O demônio M tem 2 bits de entropia, Y tem dois bits de entropia, e sua informação mútua é 2 bits, então (M,Y) tem um total de 2 + 2 - 2 = 2 bits de entropia. O processo físico apenas transforma a "frieza" (entropia negativa, ou negentropia) da informação mútua para tornar a água real fria - depois, M tem 2 bits de entropia, Y tem 0 bits de entropia, e a informação mútua é 0. Nada de errado com isso!

E não me diga que o conhecimento é "subjetivo". O conhecimento precisa ser representado em um cérebro, e isso o torna tão físico quanto qualquer outra coisa. Para M representar fisicamente uma imagem precisa do estado de Y, é preciso que o estado físico de M se correlacione com o estado de Y. Você pode tirar vantagem termodinâmica disso - é chamado de motor de Szilárd.

Ou como E. T. Jaynes colocou, "O velho ditado 'conhecimento é poder' é uma verdade muito convincente, tanto nas relações humanas quanto na termodinâmica."

E, inversamente, um subsistema não pode aumentar em informação mútua com outro subsistema sem (a) interagir com ele e (b) realizar trabalho termodinâmico.

Caso contrário, você poderia construir um Demônio de Maxwell e violar a Segunda Lei da Termodinâmica - o que, por sua vez, violaria o Teorema de Liouville - algo proibido no modelo padrão da física.

Ou seja: **Para formar crenças precisas sobre algo, é fundamental observá-lo diretamente**. Este processo é tanto físico quanto real: qualquer mente racional "trabalha" no sentido termodinâmico, não apenas no sentido de esforço mental.

(Algumas vezes, é dito que o apagamento de bits para preparar a próxima observação é o que requer trabalho termodinâmico, mas essa distinção é apenas uma questão de palavras e perspectiva; a matemática é inequívoca.).

(Descobrir "verdades" lógicas é uma complicação que não considerarei por enquanto, em parte porque eu mesmo continuo elaborando o formalismo exato. Na termodinâmica, o conhecimento de verdades lógicas não conta como negentropia; como seria esperado, já que um computador reversível pode calcular verdades lógicas a um custo arbitrariamente baixo. Tudo isso que eu disse é verdadeiro para os logicamente oniscientes: qualquer mente menor será necessariamente menos eficiente.)

"Formar crenças precisas requer uma quantidade correspondente de evidências" é uma verdade muito convincente tanto nas relações humanas quanto na termodinâmica: se a fé cega realmente funcionasse como método de investigação, você poderia transformar água morna em eletricidade e cubos de gelo. Bastaria construir um Demônio de Maxwell que tivesse fé cega nas velocidades das moléculas.

Os motores de cognição não são tão diferentes dos motores térmicos, embora manipulem a entropia de forma mais sutil do que a queima de gasolina. Por exemplo, enquanto um motor de cognição não é perfeitamente eficiente, ele deve irradiar calor residual, assim como um motor de carro ou um refrigerador.

A "racionalidade fria" é verdadeira em um sentido que os roteiristas de Hollywood nunca sonharam (e falsa no sentido que eles sonharam).

Então, a menos que você possa me dizer qual etapa específica em seu argumento viola as leis da física ao lhe dar conhecimento verdadeiro do não visto, não espere que eu acredite que um argumento grande, elaborado e inteligente possa fazê-lo também.

187 — Crenças do Movimento Perpétuo

O ensaio anterior concluiu:

Para desenvolver crenças precisas sobre algo, é essencial observá-lo diretamente. Esse processo é físico e real, pois a mente racional funciona conforme as leis da termodinâmica, não apenas com o esforço mental. Portanto, a menos que você possa identificar um passo específico em seu argumento que viole as leis da física, o que lhe concederia um conhecimento genuíno do invisível, não espere que eu acredite que um argumento inteligente e elaborado também possa fazer isso.

Uma das principais lições da analogia matemática entre termodinâmica e cognição é que as restrições da probabilidade são inexoráveis; a probabilidade pode ser um "estado subjetivo de crença", mas as leis da probabilidade são mais rígidas do que o aço.

As pessoas são ensinadas no sistema escolar tradicional que o professor lhes diz certas coisas, e elas devem acreditar e recitar essas informações; no entanto, se um simples aluno sugere uma crença, não é obrigatório obedecer. Elas traçam uma linha entre a crença e a autoridade, considerando uma crença como uma ordem que deve ser seguida, mas uma crença probabilística como uma mera sugestão.

Elas olham para um bilhete de loteria e dizem: "Você não pode provar que não vou ganhar, certo?" Significa: "Você pode ter calculado uma probabilidade baixa de ganhar, mas, como é uma probabilidade, é apenas uma sugestão e, no fim das contas, eu posso acreditar no que eu quiser".

Aqui está um pequeno experimento: quebre um ovo no chão. A regra que afirma que o ovo não se reconstituirá espontaneamente e voltará para sua mão é puramente probabilística. Uma sugestão, se preferir. As leis da termodinâmica são probabilísticas, então elas não podem ser consideradas leis de fato, da mesma forma que "Não matarás" é uma lei… certo?

Então, por que não ignorar a sugestão? Assim, o ovo se recomporá sozinho… certo?

Pode ser útil pensar desta forma, se você ainda tiver alguma intuição persistente de que crenças incertas não são confiáveis:

Na realidade, pode haver uma chance muito pequena de que o ovo se reconstitua espontaneamente. Mas você não pode esperar que isso aconteça. Você deve esperar que o ovo se quebre. Sua crença obrigatória é que a probabilidade de reconstituição espontânea do ovo é de aproximadamente 0. As probabilidades não são certezas, mas as leis da probabilidade são como teoremas.

Se você não acredita, tenta deixar um ovo cair no chão milhões de vezes, ignorando a termodinâmica que diz que ele não vai se remontar sozinho, aí você vai ver o que acontece. As probabilidades podem ser só uma coisa em que se acredita, mas as leis que as controlam são mais fortes que o aço. Uma vez, conheci uma pessoa que achava que tinha inventado um sistema de rodas e engrenagens que gerava impulso sem reação. Ela tinha uma planilha do Excel que provava isso, mas, claro, não podia nos mostrar porque ainda estava trabalhando nela. Na mecânica clássica, quebrar a lei da conservação do momento é impossível. Então, qualquer planilha do Excel que siga as regras da mecânica clássica tem que mostrar que não existe impulso sem reação, a menos que sua máquina seja tão complexa que você tenha errado nos cálculos.

É o mesmo quando racionalistas semi-treinados, ou pouco treinados, abandonam sua arte e tentam acreditar em algo sem provas, apenas dessa vez, eles geralmente constroem vastos edifícios de justificativas, confundindo-se apenas o suficiente para esconder os truques de mágica que usam.

Descobrir onde a "mágica" ocorre pode ser doloroso, porque quando você começa a fazer perguntas, toda a estrutura do argumento deles começa a mudar e se contorcer. Mas chega um momento em que uma chance bem pequena vira uma chance bem grande, e aí as pessoas tentam acreditar sem nenhuma evidência, entram no desconhecido pensando "ninguém pode provar que eu estou errado".

Seus passos naturalmente se movem pelo terreno incerto, porque tem muito mais terreno incerto do que terreno sólido no mundo das possibilidades. Mesmo assim, tem uma quantidade (bem pequena, exponencialmente pequena) de solo na possibilidade e uma chance (bem pequena, exponencialmente pequena) de acertar por sorte, então talvez desta vez você acerte o lugar certo. É só uma probabilidade, então deve ser só uma sugestão.

O estado exato de um copo com água fervendo pode ser desconhecido para você - na verdade, é a sua ignorância sobre o estado exato que faz com que a energia cinética das moléculas se transforme em "calor", em vez de trabalho a ser extraído, como o momento de um volante girando. Assim, a água pode resfriar sua mão em vez de esquentá-la, com probabilidade diferente de zero.

Se você decidir ignorar as leis da termodinâmica e colocar a mão na água fervente, vai se queimar.

"Mas você não sabe disso!"

Eu não sei com certeza, mas é obrigatório que eu espere que isso aconteça. Probabilidades não são verdades lógicas, mas as leis da probabilidade são.

"Mas e se eu adivinhar o estado da água fervente e acertar?"

A chance de você adivinhar corretamente por sorte é ainda menor do que a chance da água fervente resfriar sua mão por acaso.

"Mas você não pode provar que eu não vou adivinhar corretamente."

Posso (e devo) atribuir uma probabilidade extremamente baixa para isso.

"Mas isso não é o mesmo que certeza."

Ei, quem sabe se você adicionar rodas e engrenagens suficientes ao seu argumento, ele não transformará água morna em eletricidade e gelo! Quero dizer, você não verá mais o porquê de isso não ser possível.

"Certo! Eu não entendo o porquê disso não ser possível! Então talvez seja!"

Outra engrenagem? Isso só deixa sua máquina menos eficiente. Ela não era uma máquina de movimento perpétuo antes, e cada engrenagem extra que você adiciona a torna ainda menos eficiente. Cada detalhe extra no seu argumento diminui necessariamente a probabilidade conjunta. A probabilidade de você ter violado a Segunda Lei da Termodinâmica sem saber exatamente como, adivinhando o estado exato da água fervente sem evidências, de modo que você possa colocar o dedo nela sem se queimar, é, necessariamente, ainda menor do que você colocar o dedo na água fervente e não se queimar.

Digo isso porque as pessoas constroem esses edifícios de argumentos enormes baseados em crenças sem evidências. É preciso aprender a ver isso como análogo a todas as rodas e engrenagens que o sujeito adicionou à sua máquina de propulsão sem reação, até que finalmente reuniu complicações suficientes para errar nos cálculos em sua planilha de Excel.

188 — Procurando pela estrutura de Bayes

> Os capacetes gnômicos não deveriam funcionar. Sua própria construção parece desafiar a natureza das leis da taumaturgia. De fato, eles são impossíveis. Como muitos produtos das mentes gnômicas, eles incluem um grande número de sinos e apitos, mas pouca substância. Os capacetes que realmente funcionam geralmente escondem um capacete menor em seu interior, cuidadosamente disfarçado para parecer comum e não essencial.
>
> — Spelljammer: Cenário de Campanha (*Advanced Dungeons and Dragons*)[14]

Aprendemos anteriormente que conhecimento implica uma troca de informações entre a sua mente e o mundo ao redor, e vimos que essa troca de informações é negentropia, em um sentido muito físico: se você sabe onde as moléculas estão e como elas estão se movendo, pode transformar calor em trabalho usando um motor do tipo Demônio de Maxwell/Szilárd.

Também discutimos que criar crenças verdadeiras sem evidências é tão improvável quanto um copo de água quente se transformar espontaneamente em cubos de gelo e eletricidade. A racionalidade exige "trabalho" no sentido termodinâmico, não apenas esforço mental. As mentes precisam dissipar calor se não forem perfeitamente eficientes. Esse trabalho cognitivo é governado pela teoria da probabilidade, da qual a termodinâmica é um caso especial. (A mecânica estatística é um caso especial da estatística.)

Se você visse uma máquina girando uma roda sem parar, aparentemente sem estar conectada a uma tomada ou fonte de energia visível, você procuraria por uma bateria escondida ou uma fonte de energia de transmissão próxima. Alguma coisa que explicasse o trabalho sendo feito sem violar as leis da física.

Se uma mente está chegando a crenças verdadeiras, e assumimos que a Segunda Lei da Termodinâmica não foi violada, essa mente deve estar realizando algo vagamente bayesiano, pelo menos algum processo com alguma forma de estrutura Bayesiana em algum lugar, ou isso não funcionaria.

No início, no tempo T = 0, uma mente não possui informações mútuas com um subsistema S em seu ambiente. No tempo T = 1, a mente possui 10 bits de informações mútuas com S. Em algum momento intermediário, a mente deve ter encontrado evidências - de acordo com a definição bayesiana de evidência, porque todas as evidências bayesianas são informações mútuas e todas as informações mútuas são evidências bayesianas, são apenas maneiras diferentes de ver as coisas - e processado pelo menos algumas delas, embora ineficientemente, na direção correta, de acordo com Bayes em pelo menos algumas ocasiões. A mente deve ter se movido em harmonia com Bayes pelo menos um pouco, em algum ponto ao longo da linha - ou isso, ou violou a Segunda Lei da Termodinâmica ao criar informações mútuas a partir do nada.

Na verdade, qualquer parte de um processo cognitivo que nos ajude a descobrir a verdade deve ter pelo menos um pouco de estrutura bayesiana - deve se harmonizar com Bayes, em algum momento ou outro - deve estar parcialmente em conformidade com o fluxo bayesiano, por mais barulhento que seja - apesar de muitos sinos e assobios disfarçadores - mesmo que essa estrutura bayesiana só seja aparente no contexto dos processos circundantes. Se não, não ajudará muito.

14 NT. RPG. *Spelljammer* é um cenário de campanha para *Advanced Dungeons & Dragons* que combina fantasia medieval e aventuras espaciais, permitindo que personagens viajem em navios mágicos pelo "espaço selvagem". Lançado no início dos anos 1990 pela TSR, ele interliga mundos de AD&D, como *Forgotten Realms* e *Dragonlance*, por meio de portais e "helmets" mágicos que impulsionam as naves.

Como os filósofos refletiam sobre a natureza das palavras! Toda a tinta gasta em definições verdadeiras de palavras, e no verdadeiro significado das definições, e no verdadeiro significado do significado! Que coleções de engrenagens e rodas eles construíram em suas explicações! E, durante todo esse tempo, era uma forma disfarçada de inferência bayesiana!

Fiquei um pouco decepcionado de que ninguém na plateia tenha pulado e dito: "Sim! Sim, é isso! Claro! Sempre foi Bayes!".

Mas talvez não seja tão empolgante ver algo que não parece bayesiano à primeira vista, revelado como Bayes disfarçado inteligentemente, se: (a) você não desvenda o mistério por si mesmo, mas lê sobre alguém mais fazendo isso (Newton se divertiu mais do que a maioria dos estudantes estudando cálculo), e (b) você não percebe que procurar pela estrutura bayesiana oculta é uma busca enorme, difícil e onipresente, como procurar pelo Santo Graal.

É uma busca diferente para cada aspecto da cognição, mas o Graal sempre acaba sendo o mesmo. Tem que ser o Graal certo, no entanto - e o Graal completo, sem nenhuma parte faltando - e assim, toda vez que você tem que embarcar na busca por uma resposta completa, não importa a forma que ela possa assumir, em vez de tentar construir artificialmente argumentos *Graalistas* vagos e imprecisos. E é sempre o mesmo Santo Graal que você encontra no final.

Já me apontaram que eu poderia estar perdendo alguns dos meus leitores com os ensaios longos, porque eu não "deixei claro para onde eu estava indo"…

… Mas não é tão fácil apenas dizer às pessoas para onde você está indo, quando você está indo para um lugar assim.

Não é muito útil apenas saber que uma forma de cognição é bayesiana, se você não sabe como ela é bayesiana. Se você não consegue ver o fluxo detalhado de probabilidade, você não tem nada além de uma senha - ou, um pouco mais caridosamente, uma dica sobre a forma que uma resposta poderia ter; mas certamente não uma resposta. É por isso que existe uma Grande Busca pela Estrutura Bayesiana Oculta, em vez de simplesmente dizer "Bayes!" e achar que já fez tudo o que precisava. A estrutura bayesiana pode estar enterrada sob todo tipo de disfarces, escondida atrás de arbustos de rodas e engrenagens, obscurecida por sinos e assobios.

A maneira como você começa a entender a Busca pelo Santo Bayes, é aprender sobre o fenômeno cognitivo XYZ, que parece realmente útil - e há um grupo de filósofos que vem discutindo sua natureza há séculos e continua a ser objeto de debate - e há um grupo de cientistas de IA tentando replicar o fenômeno XYZ usando computadores, mas também não chegaram a um consenso sobre a filosofia.

E - que estranho! Esse fenômeno cognitivo não parecia nada bayesiano à primeira vista, no entanto, há uma estrutura subjacente não óbvia com uma interpretação bayesiana - mas espere, ainda há algum trabalho útil sendo feito que não pode ser explicado em termos bayesianos - não, espere, isso também é bayesiano - oh meu deus, esse processo cognitivo completamente diferente, que também não parecia bayesiano à primeira vista, também tem uma estrutura bayesiana - espere, essas partes não bayesianas estão realmente fazendo algo?

- Sim: Essas coisas são muito bayesianas!
- Não: Puxa vida, esse projeto é uma porcaria. Eu poderia comer um balde inteiro de aminoácidos e vomitar uma arquitetura cerebral melhor do que essa.

Depois que isso acontece com você algumas vezes, você começa a pegar o ritmo. É sobre isso que estou falando aqui, do ritmo.

Tentar explicar o ritmo é similar a tentar dançar sobre a arquitetura.

Isso me colocou em uma situação um pouco complicada ao tentar antecipar meu destino. Com base na minha experiência, se eu dissesse: "Bayes é o segredo do universo", algumas pessoas poderiam dizer: "Sim! Bayes é o segredo do universo!"; e outras irão bufar e dizer: "Como você é limitado; veja todos esses

outros métodos ad-hoc, mas incrivelmente úteis, como a regressão linear regularizada, que tenho em minha caixa de ferramentas."

Eu esperava que, usando um exemplo específico em mãos de "algo que não parece muito bayesiano no começo, mas acaba sendo bayesiano no final das contas" - e explicando a diferença entre senhas e conhecimento, e entre ferramentas e leis, eu pudesse transmitir um pouco do ritmo que pode ser entendido sem ter que embarcar pessoalmente na busca.

Claro, esse não é o Segredo Completo da Conspiração Bayesiana, mas é tudo que posso transmitir agora. Além disso, o segredo completo é conhecido apenas pelo Conselho Bayes, e se eu contasse para você, teria que contratar você.

Para ver através da "adhocracia" superfícial de um processo cognitivo, até a estrutura bayesiana subjacente - para perceber os fluxos de probabilidade e saber como, não apenas saber que, essa cognição também é bayesiana - como sempre é - como sempre deve ser - para ser capaz de sentira Força subjacente a toda cognição - isso é a Visão Bayesiana.

> "... E a Rainha de Kashfa vê com o Olho da Serpente."
>
> "Eu não sei se ela vê com ele", disse eu. "Ela continua se recuperando da operação. Mas essa é uma ideia interessante. Se ela pudesse ver com ele, o que ela poderia contemplar?"
>
> "As linhas claras e frias da eternidade, eu diria. Abaixo de todas as Sombras[15]."
>
> —Roger Zelazny, *Prince of Chaos* (Príncipe do Caos) [1]

Referências

[1] Roger Zelazny, Prince of Chaos (Thorndike Press, 2001). ://www.jstor.org/stable/2183914.

15 NT. Texto original em inglês. "I don't know that she sees with it," I said. "She's still recovering from the operation. But that's an interesting thought. If she could see with it, what might she behold?" "The clear, cold lines of eternity, I daresay. Beneath all Shadow."

Parte P — Reducionismo 101

189 — Desconstruindo a pergunta

"Se uma árvore cai na floresta, mas ninguém ouve, ela faz barulho?"

Eu não respondi a essa pergunta. Eu não escolhi uma posição "Sim!" ou "não!" para defender. Em vez disso, desconstruí o algoritmo humano para processar palavras, chegando até a esboçar uma ilustração de uma rede neural. No final, espero, não restou nenhuma pergunta, nem mesmo o sentimento de uma pergunta. Muitos filósofos, especialmente filósofos amadores e filósofos antigos, compartilham um instinto perigoso: se você fizer uma pergunta, eles tentarão respondê-la.

Por exemplo, digamos: "Temos livre arbítrio?"

O instinto perigoso da filosofia é organizar os argumentos a favor, organizar os argumentos contra, avaliá-los e publicá-los em uma revista prestigiosa de filosofia e, finalmente, concluir: "Sim, devemos ter livre arbítrio" ou "Não, não podemos ter livre arbítrio."

Certos filósofos têm a sabedoria de lembrar o aviso que a maioria das brigas filosóficas são, na verdade, disputas sobre o significado de uma palavra ou confusões causadas pelo uso de significados diferentes para a mesma palavra em lugares diferentes.

Então, eles tentam definir com muita precisão o que querem dizer com "livre arbítrio" e depois perguntam de novo: "Nós temos livre arbítrio? Sim ou não?"

Um filósofo ainda mais sábio pode suspeitar que a confusão sobre "livre-arbítrio" mostra que a própria ideia é falha. Assim, eles seguem o caminho racionalista tradicional : eles argumentam que o "livre arbítrio" é inerentemente autocontraditório ou sem sentido porque não tem consequências testáveis. E depois eles publicam essas observações devastadoras em uma revista de filosofia de prestígio.

Mas provar que você está confuso pode não fazer você se sentir menos confuso. Provar que uma pergunta não tem sentido pode não ajudar mais do que respondê-la.

O instinto do filósofo é encontrar a posição mais defendível, publicar e seguir em frente. Mas a visão "ingênua", a visão instintiva, é um fato sobre a psicologia humana. Você pode provar que o livre-arbítrio é impossível até o Sol esfriar, mas isso deixa um fato inexplicável da ciência cognitiva: se o livre-arbítrio não existe, o que está acontecendo na cabeça de um ser humano que pensa que existe? Esta não é uma pergunta retórica!

É um fato sobre a psicologia humana que as pessoas pensam que têm livre-arbítrio. Encontrar uma posição filosófica mais defensável não muda ou explica esse fato psicológico. A filosofia pode levar você a rejeitar o conceito, mas rejeitar um conceito não é o mesmo que entender os algoritmos cognitivos por trás dele.

Você pode considerar a Disputa Padrão sobre "Se uma árvore cai na floresta e ninguém a ouve, ela faz barulho?", e poderia fazer o que o Racionalista Tradicional faz: observar que os dois lados não discordam em nenhum ponto de experiência antecipada e declarar triunfantemente o argumento como sem sentido. Acontece que Isso está correto neste caso específico; mas, como uma questão de ciência cognitiva, por que os debatedores cometeram esse erro em primeiro lugar?

A ideia central do programa de heurísticas e vieses é que os erros que cometemos muitas vezes revelam muito mais sobre nossos algoritmos cognitivos subjacentes do que nossas respostas corretas. Então (perguntei a mim mesmo, certa vez), que tipo de design mental corresponde ao erro de discutir sobre árvores caindo em florestas desertas?

Os algoritmos cognitivos que usamos são como percebemos o mundo. E esses algoritmos cognitivos podem não ter uma correspondência de um para um com a realidade. Pode haver coisas em nossas mentes que distorcem o mundo.

Por exemplo, pode haver uma unidade pendente no centro de uma rede neural que não corresponde a nada real, nem a nenhuma propriedade real de qualquer coisa real que exista em qualquer lugar do mundo real. Essa unidade pendente geralmente é útil como um atalho na computação, e é por isso que as temos. (Falando metaforicamente, claro. A neurobiologia humana é certamente muito mais complexa.)

Essa unidade pendente parece uma questão não resolvida, mesmo depois de todas as perguntas passíveis de resposta serem respondidas. Não importa o quanto alguém tente te provar que nenhuma diferença de experiência antecipada depende da pergunta, você fica se perguntando: “Mas a árvore caindo realmente faz barulho ou não?”

Ao compreender detalhadamente como seu cérebro cria a sensação de pergunta - quando perceber que o sentimento de pergunta não respondida corresponde uma unidade central ilusória que visa saber se deve disparar, mesmo depois de todas as unidades de borda estarem fixadas em valores conhecidos - ou melhor ainda, quando entender o funcionamento técnico do Naïve Bayes—então estará tudo bem. Assim não há nenhum sentimento persistente de confusão, nenhum sentimento vago de insatisfação.

Se houver algum sentimento persistente de uma pergunta sem resposta, ou de ter sido convencido a fazer algo rapidamente, isso é um sinal de que você não dissolveu a pergunta. Uma vaga insatisfação deveria ser tanto um aviso quanto um grito. Realmente dissolver a questão não deixa nada para trás.

Uma refutação estrondosa e triunfante do livre-arbítrio, uma prova absolutamente indiscutível de que o livre-arbítrio não pode existir, pode parecer muito satisfatória - uma grande torcida para o time da casa. No entanto, você pode não perceber que, do ponto de vista da ciência cognitiva, não tem uma explicação descritiva completa e satisfatória de como cada sensação intuitiva surge, ponto por ponto.

Você pode até evitar admitir essa ignorância sobre a ciência cognitiva, pois isso pareceria um gol contra para sua equipe. Em meio de esmagar todas as crenças sobre o livre-arbítrio, admitir que deixou algo sem explicação pareceria uma concessão ao lado oposto.

E assim, talvez, você apresente um argumento psicológico evolucionista, sugerindo que os caçadores-coletores que acreditavam no livre-arbítrio eram mais propensos a ter uma visão positiva da vida e, assim, reproduzir mais - para dar um exemplo de uma explicação completamente falsa. Se fizer isso, estará argumentando que o cérebro gera uma ilusão de livre-arbítrio - mas sem explicar como. Você está tentando descartar a oposição desconstruindo seus motivos, mas na história que conta, a ilusão do livre-arbítrio permanece como um fato bruto. Você não desmontou a ilusão para ver suas rodas e engrenagens.

Imagine que, na Disputa Padrão sobre uma árvore caindo em uma floresta deserta, você primeiro prova que não existe diferença de antecipação e, em seguida, faça a hipótese: “Mas talvez as pessoas que disseram que os argumentos não tinham sentido foram vistas como tendo cedido, e, portanto perderam posição social, de modo que agora temos o instinto de discutir sobre os significados das palavras”. Isso é argumentar ou explicar por que existe uma confusão. Agora observe a estrutura da rede neural em “Feel the Meaning”. Isso está explicando como, desmontando a confusão em pedaços menores que não são confusos. Vê a diferença?

Desenvolver boas hipóteses sobre algoritmos cognitivos (ou mesmo hipóteses que se sustentem por meio segundo) é significativamente mais desafiador do que simplesmente refutar uma confusão filosófica. De fato, é uma habilidade distinta. Reconhecer isso permite que você diga com tranquilidade: “Sei que sua afirmação é falsa e posso provar isso. Mas não posso escrever um fluxograma que mostre como seu cérebro

comete o erro, portanto, ainda não terminei e continuarei investigando."

Digo tudo isso porque, às vezes, parece-me que pelo menos 20% da eficácia de um racionalista habilidoso no mundo real advém do fato de não parar muito cedo. Se você continuar fazendo perguntas, acabará chegando ao seu destino. Se você decidir muito cedo que encontrou uma resposta, não chegará.

O desafio, acima de tudo, é perceber quando você está confuso - mesmo que pareça apenas um pouquinho de confusão - e mesmo que haja alguém à sua frente insistindo que os seres humanos têm livre-arbítrio e sorrindo para você, e que o fato de você não saber exatamente como os algoritmos cognitivos funcionam não tenha nada a ver com a loucura gritante da posição dessa pessoa...

Mas quando você consegue detalhar o algoritmo cognitivo com detalhes suficientes para traçar o processo de pensamento, passo a passo, e descrever como cada intuição surge - decompor a confusão em elementos menores que não sejam confusas - então você terá terminado.

Portanto, esteja avisado de que você pode acreditar que terminou, quando tudo o que você tem é uma mera refutação triunfante de um erro.

Mas quando você realmente terminar, saberá que terminou. Dissolver a pergunta é uma sensação inconfundível - uma vez que você a experimenta e, tendo-a experimentado, decide não ser enganado novamente. Aqueles que sonham não sabem que estão sonhando, mas quando você acorda, sabe que está acordado.

Isso quer dizer que: Quando terminar, você saberá que terminou, mas infelizmente a implicação inversa não se sustenta.

Portanto, aqui está seu problema de lição de casa: que tipo de algoritmo cognitivo, sentido de dentro, geraria o debate observado sobre "livre-arbítrio"?

Sua tarefa não é discutir se as pessoas têm ou não livre-arbítrio.

Sua tarefa não é argumentar que o livre-arbítrio é compatível com o determinismo ou não.

Sua tarefa não é argumentar que a questão é mal colocada, ou que o conceito é autocontraditório, ou que não tem consequências testáveis.

Não se pede que você invente uma explicação evolucionária de como as pessoas que acreditavam no livre-arbítrio teriam se reproduzido; nem um relato de como o conceito de livre-arbítrio parece suspeitamente congruente com o viés X. Essas são meras tentativas de explicar por que as pessoas acreditam no "livre-arbítrio", não de explicar como.

Sua tarefa de casa é escrever um traço de pilha dos algoritmos internos da mente humana à medida que eles produzem as intuições que alimentam todo o maldito argumento filosófico.

Esse é um dos primeiros desafios reais que experimentei como aspirante a racionalista, em uma época passada. Um dos enigmas mais fáceis, relativamente falando. Que ele lhe sirva da mesma forma.

190 — Perguntas erradas

Quando sua mente colide com a realidade, ela pode gerar perguntas equivocadas - perguntas que não podem ser respondidas em seus próprios termos, mas apenas dissolvidas pela compreensão do algoritmo cognitivo que gera a percepção de uma pergunta.

Um bom sinal de que você está lidando com uma "pergunta errada" é quando não consegue nem imaginar algum estado concreto e específico de como o mundo é que responderia à pergunta. Pode até parecer impossível responder.

Veja a Disputa de Definição Padrão, por exemplo, sobre a queda de uma árvore em uma floresta deserta. Existe algum jeito do mundo ser - algum estado das coisas - que faça com que a palavra "som" signifique apenas vibrações acústicas ou apenas experiências auditivas?

("Ora, sim", diz aquele que diz "é o estado das coisas em que 'som' significa vibrações acústicas." Portanto, use Tabu na palavra "significa", "representa" e todos os sinônimos semelhantes, e descreva novamente: Que maneira o mundo pode ser, que estado de coisas faria com que um lado estivesse certo e o outro errado?)

Ou, se isso parecer muito fácil, considere o livre-arbítrio: qual estado concreto de coisas, seja na física determinística ou na física com um componente aleatório de lançamento de dados, poderia corresponder a ter livre-arbítrio?

E se isso parecer muito fácil, pergunte-se: "Por que algo existe?" e depois me diga como seria uma resposta satisfatória para essa pergunta.

E não, não sei a resposta para essa última pergunta. Mas posso supor uma coisa, com base em experiências anteriores com perguntas sem resposta. A resposta não envolverá uma Causa Primeira grande e triunfante. A pergunta se dissipará como resultado de algum insight sobre como meus algoritmos mentais estão distorcidos em relação à realidade, após esse insight, compreenderei como a própria pergunta estava errada desde o início - como a pergunta em si assumiu a falácia, continha a distorção.

O mistério existe na mente, não na realidade. Se sou ignorante sobre um fenômeno, isso é um fato sobre o meu estado mental, não um fato sobre o fenômeno em si. Ainda mais se parecer que não há resposta possível: a confusão existe no mapa, não no território. Perguntas sem resposta não marcam lugares onde a magia entra no universo. Elas marcam lugares onde sua mente se desvia da realidade.

Essas questões devem ser dissolvidas. Coisas ruins acontecem quando tentamos respondê-las. Isso invariavelmente leva à pior forma de Resposta Misteriosa para uma Pergunta Misteriosa: aquela em que apresentamos argumentos aparentemente sólidos para nossa Resposta Misteriosa, mas a "resposta" não permite fazer novas previsões, nem mesmo em retrospecto, e o fenômeno ainda mantém a mesma inexplicabilidade sagrada que tinha no início.

Eu poderia dizer que a resposta para o enigma da Causa Primeira é que nada existe de verdade - que todo o conceito de "existência é falso. Mas se você acreditasse sinceramente nisso, ficaria menos confuso? Eu também não.

Mas o que há de maravilhoso nas perguntas sem resposta é que ela sempre têm solução, pelo menos em minha experiência.

O que passou pela mente da Rainha Elizabeth I logo pela manhã, quando ela acordou em seu quadragésimo aniversário? Como posso imaginar facilmente respostas para essa pergunta, reconheço prontamente que talvez nunca possa respondê-la de fato, pois as informações verdadeiras se perderam no tempo.

Por outro lado, a pergunta “Por que algo existe?” parece tão absolutamente impossível que posso supor que estou simplesmente confuso de uma forma ou de outra, e a verdade provavelmente não é tão complexa em um sentido absoluto e, quando a confusão se dissipar, poderei vê-la.

Isso pode parecer contraintuitivo para aqueles que nunca lidaram com uma pergunta sem resposta, mas garanto que é assim que as coisas funcionam.

A seguir: um método simples para abordar as “perguntas erradas”.

191 — Corrigindo uma pergunta errada

Quando você se depara com uma pergunta sem resposta - uma pergunta para a qual parece impossível até mesmo imaginar uma resposta - existe um truque simples que pode torná-la solucionável.

Compare:

- “Por que eu tenho livre-arbítrio?”
- “Por que eu acredito que tenho livre-arbítrio?”

A beleza da segunda pergunta é que ela tem garantia de ter uma resposta real, independentemente de existir ou não algo como livre-arbítrio. Perguntar “Por que eu tenho livre-arbítrio?” ou “Eu tenho livre-arbítrio?” leva você a pensar em detalhes minúsculos das leis da física, tão distantes do nível macroscópico que você não conseguiria nem começar a vê-los a olho nu. E você está perguntando “Por que X é o caso?” onde X pode nem ser coerente, muito menos ser o caso.

“Por que eu acredito que tenho livre-arbítrio?”, por outro lado, tem garantia de resposta. Você, de fato, acredita que tem livre-arbítrio. Essa crença parece muito mais sólida e compreensível do que a efemeridade do livre-arbítrio. E há, de fato, uma cadeia sólida de causa e efeito cognitivo levando a essa crença.

Se você já superou o livre-arbítrio, escolha um destes substitutos:

- “Por que o tempo se move para frente em vez de para trás?” versus “Por que eu acredito que o tempo se move para frente em vez de para trás?”
- “Por que nasci como eu mesmo e não como outra pessoa?” versus “Por que eu acredito que nasci como eu mesmo e não como outra pessoa?”
- “Por que eu sou consciente?” versus “Por que eu acredito que sou consciente?”
- “Por que a realidade existe?” versus “Por que eu acredito que a realidade existe?”

A beleza deste método é que ele funciona independentemente de a pergunta ser confusa ou não. Enquanto digito isso, estou usando meias. Eu poderia perguntar “Por que estou usando meias?” ou “Por que eu acredito que estou usando meias?” Digamos que eu faça a segunda pergunta. Traçando a cadeia de causalidade, encontro:

- Eu acredito que estou usando meias porque posso ver meias nos meus pés.
- Eu vejo meias nos meus pés porque minha retina está enviando sinais de meias para meu córtex visual.
- Minha retina está enviando sinais de meias porque luz em forma de meia está incidindo na minha retina.
- Luz em forma de meia incide na minha retina porque reflete das meias que estou usando.
- Reflete das meias que estou usando porque estou usando meias.
- Estou usando meias porque as coloquei.
- Coloquei meias porque acreditava que, caso contrário, meus pés ficariam frios.
- E assim por diante.

Traçando a cadeia de causalidade, passo a passo, descubro que minha crença de que estou usando meias é totalmente explicada pelo fato de que estou usando meias. Isso é correto e apropriado, já que não se

pode obter informações sobre algo sem interagir com ele.

Por outro lado, se vejo uma miragem de um lago no deserto, a explicação causal correta da minha visão não envolve o fato de haver qualquer lago real no deserto. Neste caso, minha crença no lago não é apenas explicada, mas eliminada.

Mas de qualquer forma, a crença em si é um fenômeno real ocorrendo no universo real - eventos psicológicos são eventos - e sua história causal pode ser rastreada.

"Por que há um lago no meio do deserto?" pode falhar se não houver lago a ser explicado. Mas "Por que eu percebo um lago no meio do deserto?" sempre tem uma explicação causal, garantidamente.

Talvez alguém veja uma oportunidade de ser esperto e diga: "Ok. Eu acredito no livre-arbítrio porque tenho livre-arbítrio. Pronto, acabei." Claro que não é tão fácil assim.

Minha percepção de meias nos meus pés é um evento no córtex visual. O funcionamento do córtex visual pode ser investigado pela ciência cognitiva, caso seja confuso.

Minha retina recebendo luz não é um procedimento místico de detecção, um detector mágico de meias que se acende na presença de meias sem nenhum motivo explicável; existem mecanismos que podem ser entendidos em termos de biologia. Os fótons que entram na retina podem ser entendidos em termos de óptica. A reflexão da superfície do sapato pode ser entendida em termos de eletromagnetismo e química. Meus pés ficando frios podem ser entendidos em termos de termodinâmica.

Então não é tão fácil quanto dizer: "Eu acredito que tenho livre-arbítrio porque eu o tenho - pronto, acabei!" Você precisa conseguir quebrar a cadeia causal em etapas menores e explicar as etapas em termos de elementos que não sejam confusos por si só.

A interação mecânica da minha retina com minhas meias é bastante clara e pode ser descrita em termos de componentes não confusos, como fótons e elétrons. Onde está o sensor de livre-arbítrio no seu cérebro, e como ele detecta a presença ou ausência de livre-arbítrio? Como o sensor interage com o evento detectado, e quais são os detalhes mecânicos dessa interação?

Se sua crença realmente deriva da observação válida de um fenômeno real, chegaremos eventualmente a esse fato, se começarmos a traçar a cadeia causal de trás para frente a partir da sua crença.

Se o que você está realmente vendo é sua própria confusão, traçar a cadeia de causalidade encontrará um algoritmo que funciona de forma enviesada em relação à realidade.

De qualquer forma, a pergunta tem garantia de ter uma resposta. Você até tem um ponto de partida concreto e agradável para começar a traçar - sua crença, sentada solidamente em sua mente.

A ciência cognitiva pode não parecer tão elevada e gloriosa quanto a metafísica. Mas pelo menos as questões da ciência cognitiva são solucionáveis. Encontrar uma resposta pode não ser fácil, mas pelo menos uma resposta existe.

Ah, e mais uma coisa: a ideia de que a ciência cognitiva não é tão elevada e gloriosa quanto a metafísica é simplesmente errada. Alguns leitores estão começando a perceber isso, eu espero.

192 — A falácia da projeção mental

Nos primórdios da ficção científica, os invasores alienígenas ocasionalmente sequestravam uma garota com o vestido rasgado e a carregavam com a intenção de violentá-la, como *amorosamente* (contém ironia) retratado em muitas capas antigas de revistas. Curiosamente, os alienígenas nunca iam atrás de homens com camisas rasgadas.

Será que um alienígena não-humanoide, com uma história evolutiva e psicologia evolutiva diferentes, desejaria sexualmente uma fêmea humana? Parece bastante improvável. Para dizer mínimo.

As pessoas não cometem erros como esse raciocinando deliberadamente: "Todas as mentes possíveis são provavelmente conectadas de maneira muito semelhante, portanto, um monstro de olhos esbugalhados achará as fêmeas humanas atraentes." Provavelmente o artista nem sequer pensou em questionar se um alienígena percebe as fêmeas humanas como atraentes. Em vez disso, uma fêmea humana com um vestido rasgado é sexy - inerentemente sexy, como uma propriedade intrínseca.

Aqueles que se desviaram não pensaram sobre a história evolutiva do alienígena; eles se concentraram no vestido rasgado da mulher. Se o vestido não estivesse rasgado, a mulher seria menos sexy; o monstro alienígena nem entra na equação.

Aparentemente, nós instintivamente representamos a Sensualidade como um atributo direto da estrutura de dados Mulher, **`Mulher.sensualidade`**, como **`Mulher.altura`** ou **`Mulher.peso`**.

Se seu cérebro usa essa estrutura de dados, ou algo metaforicamente similar a ela, então de dentro parece que a sensualidade é uma propriedade inerente da mulher, não uma propriedade do alienígena olhando para a mulher. Uma vez que a mulher é atraente, o monstro alienígena será atraído por ela - não é lógico?

E. T. Jaynes usou o termo Falácia da Projeção Mental para denotar o erro de projetar as propriedades da sua própria mente no mundo externo. Jaynes, como um grão-mestre tardio da Conspiração Bayesiana, estava mais preocupado com o tratamento equivocado das probabilidades como propriedades inerentes dos objetos, em vez de estados de conhecimento parcial em alguma mente particular. Mais sobre isso em breve.

Mas a Falácia da Projeção Mental se generaliza como um erro. Está no argumento sobre o verdadeiro significado da palavra som, e na capa da revista do monstro carregando uma mulher com o vestido rasgado, e na declaração de Kant de que o espaço, por sua própria natureza, é plano, e na definição de Hume de ideias a priori como aquelas "descobríveis pela mera operação do pensamento, sem dependência do que existe em qualquer lugar do universo"…

(A propósito, uma vez li uma história de ficção científica sobre um homem humano que entrou em um relacionamento sexual com uma planta alienígena senciente de frondes apropriadamente macias; descobriu que era uma planta andrônica (masculina); agonizou sobre isso por um tempo; e finalmente decidiu que não importava muito àquela altura. E em *Illegal Aliens* (Imigrantes ilegais) de Foglio e Pollotta, os humanos pousam em um planeta habitado por insetos sencientes e veem um anúncio de filme mostrando um humano carregando um inseto em um delicado vestido de chifom. Só achei que deveria mencionar isso.)

193 — A probabilidade está na mente

No ensaio anterior, falei sobre a Falácia da Projeção Mental, usando o exemplo do monstro alienígena que rapta uma garota com um vestido rasgado para violentá-la – um equívoco que atribuí à tendência do artista de pensar que a sensualidade de uma mulher é uma propriedade da própria mulher, *mulher.sensualidade*, em vez de algo que existe na mente de um observador, e provavelmente não existiria na mente de um alienígena.

O termo "Falácia da Projeção Mental" foi cunhado pelo grande mestre Bayesiano E. T. Jaynes, como parte de sua longa e árdua batalha contra os frequentistas malditos. Jaynes acreditava que as probabilidades estavam na mente, não no ambiente – que as probabilidades expressam ignorância, estados de informação parcial; e se sou ignorante sobre um fenômeno, isso é um fato sobre meu estado mental, não um fato sobre o fenômeno em si.

Não posso fazer justiça a essa antiga guerra em poucas palavras, mas o exemplo clássico do argumento é o seguinte:

Você tem uma moeda.

A moeda é viciada.

Você não sabe para qual lado ela é viciada ou o quanto ela é viciada. Alguém simplesmente lhe disse: "A moeda é viciada", e isso foi tudo o que disseram.

Esta é toda a informação que você tem, e a única informação que você tem. Você pega a moeda, lança-a para cima e a espalma.

Agora, antes de remover a mão e olhar o resultado, você está disposto a dizer que atribui uma probabilidade de 0,5 à moeda ter caído com a cara para cima?

O frequentista diz: "Não. Dizer 'probabilidade 0,5' significa que a moeda tem uma propensão inerente a cair com a cara para cima com a mesma frequência que com a coroa para cima, de modo que se jogássemos a moeda infinitas vezes, a proporção de caras para coroas se aproximaria de 1:1. Mas sabemos que a moeda é viciada, então ela pode ter qualquer probabilidade de cair com a cara para cima, exceto 0,5".

O Bayesiano diz: "A incerteza existe no mapa, não no território. No mundo real, a moeda ou caiu com a cara para cima, ou caiu com a coroa para cima. Qualquer conversa sobre 'probabilidade' deve se referir à informação que tenho sobre a moeda – meu estado de ignorância parcial e conhecimento parcial – não apenas à moeda em si. Além disso, tenho vários teoremas mostrando que se eu não tratar meu conhecimento parcial de determinada maneira, farei apostas estúpidas. Se eu tiver que planejar, planejarei para um estado de incerteza de 50/50, onde não considero os resultados condicionais à cara mais importantes em minha mente do que os resultados condicionais à coroa. Você pode chamar esse número do que quiser, mas ele precisa obedecer às leis da probabilidade, sob pena de estupidez. Portanto, não tenho a menor hesitação em chamar minha ponderação de resultados de probabilidade".

Concordo com os Bayesianos. Você pode ter notado isso em mim.

Mesmo antes de uma moeda justa ser lançada, a noção de que ela tem uma probabilidade inerente

de 50% de cair com a cara para cima pode estar simplesmente errada. Talvez você esteja segurando a moeda de tal maneira que ela esteja praticamente garantida a cair com a cara ou coroa para cima, dada a força com que você a lança e as correntes de ar ao seu redor. Mas, se você não sabe para qual lado a moeda está viciada nesta ocasião específica, e daí?

Acredito que houve um processo judicial em que alguém alegou que a loteria do draft era injusta, porque os papéis com os nomes não estavam sendo misturados o suficiente; e o juiz respondeu: "Para quem isso é injusto?"

Para tornar o experimento do lançamento da moeda repetível, como os frequentistas costumam exigir, poderíamos construir um lançador de moedas automatizado e verificar se os resultados são 50% cara e 50% coroa. Mas talvez um robô com olhos extra-sensíveis e um bom conhecimento de física, observando o lançador automático se preparar para lançar, pudesse prever a queda da moeda com antecedência – não com certeza, mas com 90% de precisão. Então, qual seria a probabilidade real?

Não existe "probabilidade real". O robô tem um estado de informação parcial. Você tem um estado de informação parcial diferente. A moeda em si não tem mente e não atribui probabilidade a nada; ela simplesmente gira no ar, gira algumas vezes, ricocheteia em algumas moléculas de ar e cai com a cara ou coroa para cima.

Então, essa é a visão Bayesiana das coisas. Agora, gostaria de apresentar alguns quebra-cabeças clássicos que derivam sua capacidade de provocar o cérebro da tendência de pensar em probabilidades como propriedades inerentes dos objetos.

Peguemos o velho clássico: Você encontra uma matemática na rua, e ela menciona que deu à luz a duas crianças em duas ocasiões diferentes. Você pergunta: "Pelo menos um de seus filhos é menino?" A matemática responde: "Sim, ele é".

Qual a probabilidade de ela ter dois meninos? Se você assumir que a probabilidade prévia de uma criança ser menino é 1/2, então a probabilidade de ela ter dois meninos, com base na informação fornecida, é 1/3. As probabilidades prévias eram: 1/4 dois meninos, 1/2 um menino e uma menina, 1/4 duas meninas. A resposta "Sim" da matemática tem probabilidade 1 nos dois primeiros casos e probabilidade 0 no terceiro. Renormalizando, ficamos com 1/3 de probabilidade de dois meninos e 2/3 de probabilidade de um menino e uma menina.

Mas suponha que, em vez disso, você tivesse perguntado: "Seu filho mais velho é um menino?" e a matemática tivesse respondido "Sim". Então a probabilidade da matemática ter dois meninos seria 1/2. Já que o filho mais velho é um menino, e o filho mais novo pode ser qualquer coisa.

Da mesma forma, se você tivesse perguntado: "Seu filho mais novo é um menino?" A probabilidade de ambos serem meninos seria, novamente, 1/2.

Agora, se pelo menos um filho é menino, deve ser o filho mais velho ou o filho mais novo que é menino. Então, como a resposta no primeiro caso pode ser diferente da resposta nos dois últimos?

Ou aqui está um problema muito semelhante: digamos que eu tenho quatro cartas, o ás de copas, o ás de espadas, o dois de copas e o dois de espadas. Eu pego duas cartas aleatoriamente. Você me pergunta: "Você está segurando pelo menos um ás?" e eu respondo "Sim". Qual a probabilidade de eu estar segurando um par de ases? É 1/5. Existem seis combinações possíveis de duas cartas, com probabilidades prévias iguais, e você acabou de eliminar a possibilidade de eu estar segurando um par de dois. Das cinco combinações restantes, apenas uma combinação é um par de ases. Então, 1/5.

Agora suponha que, em vez disso, você me perguntou: "Você está segurando o ás de espadas?" Se eu responder "Sim", a probabilidade de a outra carta ser o ás de copas é 1/3. (Você sabe que estou segurando o ás de espadas, e há três possibilidades para a outra carta, apenas uma das quais é o ás de copas.) Da mesma forma, se você me perguntar "Você está segurando o ás de copas?" e eu responder "Sim", a probabilidade de eu estar segurando um par de ases é 1/3.

Então, como pode ser que, se você me perguntar: “Você está segurando pelo menos um ás?” e eu disser “Sim”, a probabilidade de eu ter um par é 1/5? Devo estar segurando o ás de espadas ou o ás de copas, como você sabe; e de qualquer forma, a probabilidade de eu estar segurando um par de ases é 1/3.

Como isso pode ser? Calculei mal uma ou mais dessas probabilidades?

Se você quer descobrir por si mesmo, faça isso agora, porque estou prestes a revelar…

Que todos os cálculos declarados estão corretos.Quanto ao paradoxo, não há nenhum. A aparência de paradoxo surge de pensar que as probabilidades devem ser propriedades das próprias cartas. O ás que estou segurando tem que ser copas ou espadas; mas isso não significa que seu conhecimento sobre minhas cartas deva ser o mesmo que se você soubesse que eu estava segurando copas ou soubesse que eu estava segurando espadas.

Pode ajudar pensar no Teorema de Bayes:

$$P(H|E) = \frac{P(E|H)P(H)}{P(E)}$$

Esse último termo, onde você divide por P(E), é a parte em que você descarta todas as possibilidades em que foram eliminadas e renormaliza suas probabilidades sobre o que resta.

Agora, digamos que você me pergunte: “Você está segurando pelo menos um ás?” Antes de eu responder, sua probabilidade de eu dizer “Sim” deve ser 5/6.

Mas se você me perguntar “Você está segurando o ás de espadas?”, sua probabilidade prévia de eu dizer “Sim” é apenas 1/2.

Então, imediatamente você pode ver que está aprendendo algo muito diferente nos dois casos. Você estará eliminando algumas possibilidades diferentes e renormalizando usando um P(E) diferente. Se você aprender dois itens diferentes de evidência, não deve se surpreender ao acabar em dois estados diferentes de informação parcial.

Da mesma forma, se eu perguntar à matemática “Pelo menos um de seus dois filhos é menino?”, espero ouvir “Sim” com probabilidade 3/4, mas se eu perguntar “Seu filho mais velho é um menino?”, espero ouvir “Sim” com probabilidade 1/2. Portanto, não deveria ser surpreendente que eu termine em um estado diferente de conhecimento parcial, dependendo de qual das duas perguntas eu faço.

A única razão para ver um “paradoxo” é pensar como se a probabilidade de segurar um par de ases fosse uma propriedade das cartas que têm pelo menos um ás, ou uma propriedade das cartas que por acaso contêm o ás de espadas. Nesse caso, seria paradoxal que conjuntos de cartas com pelo menos um ás tenham uma probabilidade de par de 1/5, enquanto conjuntos de cartas com o ás de espadas tenham uma probabilidade de par de 1/3 e conjuntos de cartas com o ás de copas tenham uma probabilidade de par de 1/3.

Da mesma forma, se você acha que há uma probabilidade de 1/3 de serem ambos meninos, isso não faz sentido quando você olha para conjuntos de filhos onde o filho mais velho é homem, que têm uma probabilidade de 1/2 de serem ambos meninos, e conjuntos de filhos onde o filho mais novo é homem, que também têm uma probabilidade de 1/2 de serem ambos meninos. Seria como dizer: “Todas as maçãs verdes pesam meio quilo, e todas as maçãs vermelhas pesam meio quilo, e todas as maçãs que são verdes ou vermelhas pesam um quarto de quilo”.

Isso é o que acontece quando você começa a pensar como se as probabilidades estivessem nas coisas, em vez de as probabilidades serem estados de informação parcial sobre as coisas.

Probabilidades expressam incertezas, e apenas os agentes podem estar incertos. Um mapa em branco não corresponde a um território em branco. A ignorância está na mente.

194 — A citação não é o referente

Na lógica clássica, a definição operacional de identidade é que, sempre que A = B for um teorema, você pode substituir A por B em qualquer teorema onde B apareça. Por exemplo, se (2 + 2) = 4 é um teorema, e ((2 + 2) + 3) = 7 é um teorema, então (4 + 3) = 7 é um teorema.

Isso leva a um problema que geralmente é formulado nos seguintes termos: A estrela da manhã e a estrela da noite acontecem de ser o mesmo objeto, o planeta Vênus. Suponha que João saiba que a estrela da manhã e a estrela da noite são o mesmo objeto. Maria, no entanto, acredita que a estrela da manhã é o deus Lúcifer, mas a estrela da noite é a deusa Vênus. João acredita que Maria acredita que a estrela da manhã é Lúcifer. João deve, portanto (por substituição), acreditar que Maria acredita que a estrela da noite é Lúcifer?

Ou aqui está uma versão ainda mais simples do problema: A afirmação 2 + 2 = 4 é verdadeira; é um teorema que (((2 + 2) = 4) = verdadeiro). O Último Teorema de Fermat também é verdadeiro. Então: Eu acredito que 2 + 2 = 4 → Eu acredito no verdadeiro → Eu acredito no Último Teorema de Fermat.

Sim, eu sei que isso parece obviamente errado. Mas imagine alguém escrevendo um programa de raciocínio lógico usando o princípio "termos iguais sempre podem ser substituídos", e isso acontecendo com eles. Agora imagine-os escrevendo um artigo sobre como evitar que isso aconteça. Agora imagine outra pessoa discordando da solução deles. O argumento continua em andamento.

Pessoalmente, eu diria que João está cometendo um erro de tipo, como tentar subtrair 5 gramas de 20 metros. "A estrela da manhã" não é do mesmo tipo que a estrela da manhã, muito menos a mesma coisa. Crenças não são planetas.

estrela da manhã = estrela da noite

"estrela da manhã" ≠ "estrela da noite"

O problema, na minha visão, decorre da falha em impor a distinção de tipo entre crenças e coisas. O erro original foi escrever uma IA que armazena suas crenças sobre as crenças de Maria sobre "a estrela da manhã" usando a mesma representação que em suas crenças sobre a estrela da manhã.

Se Maria acredita que a "estrela da manhã" é Lúcifer, isso não significa que Maria acredita que a "estrela da noite" é Lúcifer, porque "estrela da manhã" ≠ "estrela da noite". Todo o paradoxo decorre da falha em usar aspas nos lugares apropriados.

Você pode se lembrar que esta não é a primeira vez que falo sobre impor disciplina de tipos - a última vez foi quando falei sobre o erro de confundir utilidades esperadas com utilidades. É imensamente útil manter o controle das unidades quando você está aprendendo física pela primeira vez. Pode parecer chato ficar escrevendo "cm" e "kg" e assim por diante, até que você percebe que (a) sua resposta parece ser da ordem de grandeza errada e (b) está expressa em segundos por grama quadrado. Da mesma forma, crenças são coisas diferentes de planetas. Se estamos falando sobre crenças humanas, pelo menos, então: crenças vivem em cérebros, planetas vivem no espaço. Crenças pesam alguns microgramas, planetas pesam muito mais. Planetas são maiores que crenças… mas você entendeu a ideia.

Apenas colocar aspas ao redor de "estrela da manhã" parece insuficiente para evitar que as pessoas a

confundam com a estrela da manhã, devido à similaridade visual do texto. Então, talvez uma maneira melhor de impor disciplina de tipos seria com uma codificação visivelmente diferente:

estrela da manhã = estrela da noite

13:15:18:14:9:14:7:0:19:20:1:18 ≠ 5:22:5:14:9:14:7:0:19:20:1:18

Estudar lógica matemática também pode ajudá-lo a distinguir a citação e o referente. Na lógica matemática, ⊢ P (P é um teorema) e ⊢ □⌜ P ⌝(é provável que existe uma prova codificada da sentença codificada P em algum sistema de prova codificado) são proposições muito distintas. Se você deixar cair um nível de citação na lógica matemática, é como deixar cair uma unidade métrica na física - você pode derivar resultados visivelmente ridículos, como "A velocidade da luz é 299.792.458 metros de comprimento."

Alfred Tarski[16] uma vez tentou definir o significado de "verdadeiro" usando uma família infinita de sentenças:

("A neve é branca" é verdadeiro) se e somente se (a neve é branca)

("As doninhas são verdes" é verdadeiro) se e somente se (as doninhas são verdes)

...

Quando sentenças como essas começarem a parecer significativas, você saberá que começou a distinguir entre sentenças codificadas e estados do mundo exterior.

Da mesma forma, a noção de verdade é bem diferente da noção de realidade. Dizer "verdadeiro" compara uma crença à realidade. A própria realidade não precisa ser comparada a nenhuma crença para ser real. Lembre-se disso na próxima vez que alguém afirmar que nada é verdadeiro.

16 NT. **Alfred Tarski** (1901–1983) foi um lógico e matemático polonês-americano, considerado um dos fundadores da lógica moderna. Ele desenvolveu a teoria semântica da verdade, voltada a formalizar as condições sob as quais uma sentença é considerada verdadeira. Seus estudos influenciaram fortemente a filosofia da linguagem e a teoria da lógica formal.

195 — Qualitativamente confuso

Sugiro que uma causa primária de confusão sobre a distinção entre “crença”, “verdade” e “realidade” é o pensamento qualitativo sobre crenças.

Considere a tentativa arquetípica do pós-modernista de ser esperto:

“O Sol gira em torno da Terra” é verdade para Hunga Caçador-Coletor, mas “A Terra gira em torno do Sol” é verdade para Amara Astrônoma! Sociedades diferentes têm verdades diferentes!

Não, sociedades diferentes têm crenças diferentes. Crença é de um tipo diferente da verdade; é como comparar maçãs e probabilidades.

Ah, mas não há diferença entre como você usa a palavra “crença” e como usa a palavra “verdade”! Quer você diga, “Eu acredito que ‘a neve é branca’,” ou diga,” ‘A neve é branca’ é verdade,” você está expressando a mesma opinião.

Não, essas frases significam coisas bem diferentes, o que me permite conceber a possibilidade de que minhas crenças sejam falsas.

Oh, você afirma conceber isso, mas nunca acredita nisso. Como disse Wittgenstein, “Se houvesse um verbo que significasse ‘acreditar falsamente’, ele não teria nenhuma primeira pessoa significativa no presente do indicativo.”

E é isso que quero dizer ao apontar o raciocínio qualitativo como a fonte do problema. A dicotomia entre crença e descrença, sendo binária, é confusamente similar à dicotomia entre verdade e inverdade.

Então, usemos o raciocínio quantitativo em vez disso. Suponha que eu atribua uma probabilidade de 70% à proposição de que a neve é branca. Segue-se que acho que há cerca de 70% de chance de que a frase “a neve é branca” se revele verdadeira.

Se a frase “a neve é branca” é verdadeira, minha atribuição de probabilidade de 70% à proposição também é “verdadeira”? Bem, é mais verdadeira do que teria sido se eu tivesse atribuído 60% de probabilidade, mas não tão verdadeira quanto se eu tivesse atribuído 80% de probabilidade.

Ao falar sobre a correspondência entre uma atribuição de probabilidade e a realidade, uma palavra melhor que “verdade” seria “precisão”. “Precisão” soa mais quantitativa, como um arqueiro atirando uma flecha: quão perto sua atribuição de probabilidade atingiu o centro do alvo?

Para resumir uma longa história, resulta haver uma maneira muito natural de pontuar a precisão de uma atribuição de probabilidade, em comparação com a realidade: basta tomar o logaritmo da probabilidade atribuída ao estado real das coisas.

Então, se a neve é branca, minha crença “70%: ‘a neve é branca’” pontuará -0,51 bits:

$$\log_2(0,7)= -0,51.$$

Mas e se a neve não for branca, como concedi uma probabilidade de 30% de ser o caso? Se “a neve é

branca" for falso, minha crença "30% de probabilidade: 'a neve não é branca'" pontuará -1,73 bits. Note que -1,73 < -0,51, então eu me saí pior.

Quão precisas acho que minhas próprias crenças são? Bem, minha expectativa sobre a pontuação é:

$$70\% \ (-0{,}51) + 30\% \ (-1{,}73) = -0{,}88 \text{ bits.}$$

Se a neve for branca, então minhas crenças serão mais precisas do que eu esperava; e se a neve não for branca, minhas crenças serão menos precisas do que eu esperava; mas em nenhum caso minha crença será exatamente tão precisa quanto eu esperava em média.

Tudo isso não deve ser confundido com a afirmação "Eu atribuo 70% de credibilidade a 'a neve é branca'." Posso muito bem acreditar nessa proposição com probabilidade 1 - estar bastante certo de que esta é, de fato, minha crença. Se for assim, esperarei que minha meta crença "1: 'Eu atribuo 70% de credibilidade a "a neve é branca" '" pontue 0 bits de precisão, o que é o melhor possível.

Só porque estou incerto sobre a neve, não significa que estou incerto sobre minhas crenças probabilísticas citadas. A neve está lá fora, minhas crenças estão dentro de mim. Posso estar muito menos incerto sobre quão incerto estou sobre a neve, do que estou incerto sobre a neve. (Embora crenças sobre crenças nem sempre sejam precisas.)

Contraste esta situação probabilística com o raciocínio qualitativo onde eu simplesmente acredito que a neve é branca, e acredito, que acredito que a neve é branca, e acredito que" 'a neve é branca' é verdade", e acredito que "minha crença ' "a neve é branca" é verdade' está correta", etc. Como todas as quantidades envolvidas são 1, é fácil misturá-las.

No entanto, as boas distinções do raciocínio quantitativo serão curto-circuitadas se você começar a pensar " ' "a neve é branca" com 70% de probabilidade' é verdade", o que é um erro de tipo. É uma verdade sobre você, que você acredita "70% de probabilidade: 'a neve é branca' "; mas isso não significa que a própria atribuição de probabilidade possa ser "verdadeira". A crença pontua -0,51 bits ou -1,73 bits de precisão, dependendo do estado real da realidade.

Os cognoscenti reconhecerão " ' "a neve é branca" com 70% de probabilidade' é verdade" como o erro de pensar que as probabilidades são propriedades inerentes das coisas.

De dentro, nossas crenças sobre o mundo parecem o mundo, e nossas crenças sobre nossas crenças parecem crenças. Quando você vê o mundo, está experimentando uma crença de dentro. Quando você percebe que acredita em algo, está experimentando uma crença sobre crença de dentro. Então, se suas representações internas de crença e crença sobre crença forem dissimilares, você tem menos probabilidade de confundi-las e cometer a Falácia da Projeção Mental - espero.

Quando você pensa em probabilidades, suas crenças e suas crenças sobre suas crenças, esperançosamente, não serão representadas semelhantemente o suficiente para que você confunda crença e precisão, ou confunda precisão e realidade. Quando você pensa em probabilidades sobre o mundo, suas crenças serão representadas com probabilidades no intervalo (0, 1). Ao contrário dos valores-verdade das proposições, que estão no conjunto {verdadeiro, falso}. Quanto à precisão de sua crença probabilística, você pode representá-la no intervalo $(-\infty, 0)$. Suas probabilidades sobre suas crenças tipicamente serão extremas. E as coisas em si - ora, elas são simplesmente vermelhas, ou azuis, ou pesam 9 kg, ou seja, lá o que for.

Assim, seremos menos propensos, talvez, a confundir o mapa com o território.

Esta distinção de tipo também pode nos ajudar a lembrar que a incerteza é um estado da mente. Uma moeda não é inerentemente 50% incerta sobre qual lado cairá. A moeda não é um processador de crenças e não tem informações parciais sobre si mesma. No raciocínio qualitativo, você pode criar uma crença que corresponda muito diretamente à moeda, como "A moeda cairá cara". Esta crença será verdadeira ou falsa dependendo da moeda, e haverá uma implicação transparente da verdade ou falsidade da crença para o lado visível da moeda.

Mas mesmo sob o raciocínio qualitativo, dizer que a própria moeda é “verdadeira” ou “falsa” seria um grave erro de tipo. A moeda não é uma crença. É uma moeda. O território não é o mapa.

Se uma moeda não pode ser verdadeira ou falsa, quanto menos ela pode atribuir uma probabilidade de 50% a si mesma?

196 — Pense como a realidade

Sempre que ouço alguém descrever a física quântica como “estranha” – sempre que ouço alguém lamentando os misteriosos efeitos da observação sobre o observado, ou a bizarra existência de correlações não locais, ou a incrível impossibilidade de conhecer a posição e o momento ao mesmo tempo – penso comigo mesmo: essa pessoa nunca entenderá física, não importa quantos livros leia.

A realidade existe desde muito antes de você aparecer. Não a chame de nomes desagradáveis como “bizarra” ou “incrível”. O universo propagava amplitudes complexas através do espaço de configuração por dez bilhões de anos antes da vida surgir na Terra. A física quântica não é “estranha”. Você é estranho. Você tem a ideia absolutamente bizarra de que a realidade deveria consistir em pequenas bolas de bilhar se chocando, quando, na verdade, a realidade é uma nuvem perfeitamente normal de amplitude complexa no espaço de configuração. Este é o seu problema, não da realidade, e você é quem precisa mudar.

As intuições humanas foram produzidas pela evolução, e a evolução é um hack. O mesmo processo de otimização que construiu sua retina ao contrário e depois passou o cabo óptico pelo seu campo de visão também projetou seu sistema visual para processar objetos persistentes saltando em três dimensões espaciais, porque era isso que era necessário para perseguir tigres. Mas “tigres” são generalizações superficiais e falhas – os tigres surgiram gradualmente ao longo do tempo evolutivo e não são todos absolutamente semelhantes entre si. Quando você desce ao nível fundamental, o nível no qual as leis são estáveis, globais e sem exceção, não existem tigres. Na verdade, não existem objetos persistentes saltando em três dimensões espaciais. Lide com isso.

Chamar a realidade de “estranha” o mantém em um ponto de vista já comprovadamente errôneo. A teoria da probabilidade nos diz que a surpresa é a medida de uma hipótese ruim; se um modelo é consistentemente estúpido – consistentemente se depara com eventos aos quais o modelo atribui probabilidades minúsculas – então é hora de descartar esse modelo. Um bom modelo faz a realidade parecer normal, não estranha; um bom modelo atribui alta probabilidade ao que realmente é o caso. Intuição é apenas um modelo com outro nome: intuições ruins são chocadas pela realidade, boas intuições fazem a realidade parecer natural. Você quer remodelar suas intuições para o universo parecer normal. Você quer pensar como a realidade.

Este estado final não pode ser forçado. É inútil fingir que a física quântica parece natural para você quando, na verdade, ela parece estranha. Isso é apenas negar sua confusão, não se tornar menos confuso. Mas também o impedirá de continuar pensando “Que bizarro!”. Gastar energia emocional com incredulidade desperdiça tempo que você poderia usar para se atualizar. Isso repetidamente o joga de volta na moldura do antigo ponto de vista errado. Alimenta seu senso de justa indignação com a realidade que ousa contradizê-lo.

O princípio se estende além da física. Você já se pegou dizendo algo como: “Eu simplesmente não entendo como um físico com doutorado pode acreditar em astrologia?” Bem, se você literalmente não entende, isso indica um problema com seu modelo de psicologia humana. Talvez você esteja indignado – você deseja expressar forte desaprovação moral. Mas se você literalmente não entende, então sua indignação o impede de aceitar a realidade. Não deveria ser difícil imaginar como um físico com doutorado acaba acreditando em astrologia. As pessoas compartimentalizam, basta dizer.

Agora tento evitar usar a expressão “Eu simplesmente não entendo como…” para expressar indignação. Se eu genuinamente não entendo como, então meu modelo está sendo surpreendido pelos fatos, e eu

deveria descartá-lo e encontrar um modelo melhor.

A surpresa existe no mapa, não no território. Não existem fatos surpreendentes, apenas modelos surpreendidos pelos fatos. O mesmo vale para fatos chamados por nomes tão desagradáveis como “bizarro”, “incrível”, “inacreditável”, “inesperado”, “estranho”, “anômalo” ou “estranho”. Quando você se sentir tentado por tais rótulos, pode ser sábio verificar se o suposto fato é realmente factual. Mas se o fato se confirmar, então o problema não é o fato – é você.

197 — Inversão caótica

Recentemente, tive uma conversa com alguns amigos sobre produtividade hora a hora e manutenção da força de vontade — algo com o qual lutei durante toda a minha vida.

Posso evitar fugir de um problema difícil na primeira vez que o vejo (perseverança em uma escala de tempo de segundos), e posso me ater ao mesmo problema por anos. Mas continuar trabalhando em uma escala de tempo de horas é uma batalha constante para mim. Não é preciso dizer que já li pilhas e pilhas de conselhos. O maior auxílio que obtive deles foi perceber que uma fração considerável de outros profissionais criativos tinha o mesmo problema e também não conseguiam superá-lo, não importa o quão razoáveis todos os conselhos soem.

"O que você faz quando não consegue trabalhar?" meus amigos me perguntaram. (Conversa provavelmente não precisa, esta é uma essência muito vaga.)

E eu respondi que geralmente navego em sites aleatórios ou assisto a um vídeo curto.

"Bem," eles disseram, "se você sabe que não pode trabalhar por um tempo, deveria assistir a um filme ou algo assim."

"Infelizmente," respondi, "tenho que fazer algo cujo tempo vem em unidades curtas, como navegar na Web ou assistir a vídeos curtos, porque posso me tornar capaz de trabalhar novamente a qualquer momento, e não posso prever quando—"

E então parei, porque acabara de ter uma revelação.

Sempre pensei no meu ciclo de trabalho como algo caótico, algo imprevisível. Nunca usei essas palavras, mas era assim que o tratava.

Mas aqui meus amigos pareciam estar insinuando - que pensamento estranho - que outras pessoas podiam prever quando se tornariam capazes de trabalhar novamente e estruturar seu tempo de acordo.

E me ocorreu pela primeira vez que eu poderia estar cometendo aquela maldita velha armadilha da Falácia da Projeção Mental, bem ali na minha vida cotidiana comum, em vez de em alta abstração.

Talvez não fosse que minha produtividade era incomumente caótica; talvez eu fosse apenas incomumente estúpido em relação a prevê-la.

É assim que a estupidez invertida se parece - caos. Algo difícil de lidar, difícil de compreender, difícil de adivinhar, algo com o qual você não pode fazer nada. Não é apenas uma expressão para coisas altamente abstratas como Inteligência Artificial. Pode-se aplicar na vida comum também.

E a razão pela qual não pensamos na explicação alternativa "Eu sou estúpido" não é - suspeito - que pensemos tão bem de nós mesmos. É apenas que não pensamos em nós mesmos de forma alguma. Apenas vemos uma característica caótica do ambiente.

Então agora me ocorreu que meu problema de produtividade pode não ser caos, mas minha própria estupidez.

E isso pode ou não ajudar em algo. Certamente não resolve o problema de imediato. Dizer “Sou ignorante” não o torna conhecedor.

Mas é, pelo menos, um caminho diferente do que dizer “é muito caótico”.

198 — Reducionismo

Há quase um ano, em abril de 2007, Matthew C. submeteu a seguinte sugestão para um tópico do *Overcoming Bias*:

> Como e por que o atual hegemônico filosófico reinante (materialismo reducionista) está obviamente correto […], enquanto os pontos de vista filosóficos reinantes de todas as sociedades e civilizações passadas são obviamente suspeitos[17]—

Lembro-me disso porque olhei para o pedido e o considerei legítimo, mas sabia que não poderia abordar esse tópico até que tivesse começado a sequência sobre a Falácia da Projeção Mental, o que não seria por um tempo...

Mas agora é hora de começar a abordar essa questão. E enquanto ainda não tenha chegado à questão do "materialismo", podemos agora começar com o "reducionismo".

Primeiro, que seja dito que eu realmente acredito que o "reducionismo", de acordo com o significado que darei a essa palavra, está obviamente correto; e que vão para o inferno quaisquer civilizações passadas que discordaram.

Isso parece uma declaração forte, pelo menos a primeira parte dela. A Relatividade Geral parece bem fundamentada, mas quem sabe se algum físico futuro não poderá derrubá-la?

Por outro lado, nunca voltaremos à mecânica newtoniana. A catraca da ciência gira, mas não gira ao contrário. Há casos na história da ciência em que uma teoria sofreu um ferimento ou dois e depois se recuperou; mas quando uma teoria leva tantas flechadas no peito quanto a mecânica newtoniana, ela permanece morta.

"Ao inferno com o que as civilizações passadas pensavam" parece seguro o suficiente, quando civilizações passadas acreditavam em algo que foi falsificado para a pilha de lixo da história.

E o reducionismo não é tanto uma hipótese positiva, mas a ausência de crença - em particular, a descrença em uma forma da Falácia da Projeção Mental.

Uma vez conheci um sujeito que afirmava ter experiência como artilheiro da Marinha, e ele disse: "Quando você dispara projéteis de artilharia, você tem que calcular as trajetórias usando a mecânica newtoniana. Se você calcular as trajetórias usando a relatividade, obterá a resposta errada."

E eu, e outra pessoa que estava presente, dissemos categoricamente: "Não." Acrescentei: "Você pode não conseguir calcular as trajetórias rápido o suficiente para obter as respostas a tempo - talvez seja isso que você quer dizer? Mas a resposta relativística sempre será mais precisa do que a newtoniana."

"Não," ele disse, "quero dizer que a relatividade lhe dará a resposta errada, porque as coisas que se movem na velocidade dos projéteis de artilharia são governadas pela mecânica newtoniana, não pela relati-

17 NT. Texto original em inglês. *How and why the current reigning philosophical hegemon (reductionistic materialism) is obviously correct [. . .], while the reigning philosophical viewpoints of all past societies and civilizations are obviously suspect—*

vidade."

"Se isso fosse realmente verdade," repliquei, "você poderia publicá-lo em uma revista de física e receber seu Prêmio Nobel."

A física padrão usa a mesma teoria fundamental para descrever o voo de um avião Boeing 747 e colisões no Colisor de Íons Pesados Relativísticos[18]. Tanto os núcleos quanto os aviões, de acordo com nosso entendimento, estão obedecendo à Relatividade Especial, à mecânica quântica e à cromodinâmica.

Mas usamos modelos completamente diferentes para entender a aerodinâmica de um 747 e uma colisão entre núcleos de ouro no RHIC. Um computador modelando a aerodinâmica de um 747 pode não conter um único token, um único bit de RAM, que represente um quark.

Então o 747 é feito de algo diferente de quarks? Não, você está apenas o modelando com elementos representacionais que não têm uma correspondência um-para-um com os quarks do 747. O mapa não é o território.

Por que não modelar o 747 com uma representação cromodinâmica? Porque então levaria um zilhão de anos para obter qualquer resposta do modelo. Além disso, não poderíamos armazenar o modelo em toda a memória de todos os computadores do mundo, em 2008.

Como diz o ditado, "O mapa não é o território, mas você não pode dobrar o território e colocá-lo no porta-luvas." Às vezes você precisa de um mapa menor para caber em um porta-luvas mais apertado - mas isso não muda o território. A escala de um mapa não é um fato sobre o território, é um fato sobre o mapa.

Se fosse possível construir e executar um modelo cromodinâmico do 747, ele produziria previsões precisas. Previsões melhores do que o modelo aerodinâmico, na verdade.

Para construir um modelo totalmente preciso do 747, não é necessário, em princípio, que o modelo contenha descrições explícitas de coisas como fluxo de ar e sustentação. Não precisa haver um único token, um único bit de RAM, que corresponda à posição das asas. É possível, em princípio, construir um modelo preciso do 747 que não mencione nada além de campos de partículas elementares e forças fundamentais.

"O quê?" grita o antirreducionista. "Você está me dizendo que o 747 não tem realmente asas? Posso ver as asas bem ali!"

A noção aqui é sutil. Não é apenas a noção de que um objeto pode ter descrições diferentes em níveis diferentes.

É a noção de que "ter descrições diferentes em níveis diferentes" é em si algo que você diz que pertence ao reino de Falar Sobre Mapas, não ao reino de Falar Sobre Território.

Não é que o próprio avião, as próprias leis da física, usem descrições diferentes em níveis diferentes - como pensava aquele artilheiro. Em vez disso, nós, para nossa conveniência, usamos diferentes modelos simplificados em diferentes níveis.

Se você olhasse para o modelo cromodinâmico final, aquele que continha apenas campos de partículas elementares e forças fundamentais, esse modelo conteria todos os fatos sobre fluxo de ar, sustentação e posições das asas - mas esses fatos seriam implícitos, em vez de explícitos.

Você, olhando para o modelo e pensando sobre ele, poderia descobrir onde estavam as asas. Tendo descoberto isso, haveria uma representação explícita em sua mente da posição da asa - um objeto computa-

18 NT. **Colisor de Íons Pesados Relativísticos (RHIC)**: Acelerador de partículas nos EUA (Brookhaven) que colide núcleos pesados (ex: ouro) em velocidades próximas à da luz. Objetiva recriar o *quark-gluon plasma*, estado da matéria existente logo após o Big Bang, e estudar a força nuclear forte. Seus experimentos exploram a estrutura da matéria, a evolução do universo primordial e propriedades de estrelas de nêutrons, além de avanços em física de spin e aplicações tecnológicas.

cional explícito, ali na sua RAM neural. Em sua mente.

Você poderia, de fato, deduzir todos os tipos de descrições explícitas do avião, em vários níveis, e até regras explícitas de como seus modelos em diferentes níveis interagem entre si para produzir previsões combinadas—

E a maneira como esse algoritmo se sente por dentro é que o avião pareceria ser composto de muitos níveis ao mesmo tempo, interagindo uns com os outros.

A maneira como uma crença se sente por dentro é que você parece estar olhando diretamente para a realidade. Quando realmente parece que você está olhando para uma crença, como tal, você está realmente experimentando uma crença sobre crença.

Então, quando sua mente acredita simultaneamente em descrições explícitas de muitos níveis diferentes, e acredita em regras explícitas para transitar entre níveis, como parte de um modelo combinado eficiente, parece que você está vendo um sistema feito de descrições de diferentes níveis e suas regras de interação.

Mas isso é apenas o cérebro tentando comprimir eficientemente um objeto que ele não pode nem remotamente começar a modelar em um nível fundamental. O avião é grande demais. Mesmo um átomo de hidrogênio seria grande demais. As interações quark-a-quark são insanamente intratáveis. Você não pode lidar com a verdade.

Mas a maneira como a física realmente funciona, até onde podemos dizer, é que existe apenas o nível mais básico - os campos de partículas elementares e as forças fundamentais. Você não pode lidar com a verdade crua, mas a realidade pode lidar com ela sem a menor simplificação. (Eu gostaria de saber onde a Realidade obtém seu poder de computação.)

As leis da física não contêm entidades causais adicionais distintas que correspondam à sustentação ou às asas do avião, da mesma forma que a mente de um engenheiro contém entidades cognitivas adicionais distintas que correspondem à sustentação ou às asas do avião.

Isso, como eu vejo, é a tese do reducionismo. O reducionismo não é uma crença positiva, mas sim uma descrença de que os níveis superiores de modelos simplificados em vários níveis estejam lá fora no território. Entender isso em um nível instintivo dissolve a questão de “Como você pode dizer que o avião não tem **realmente** asas, quando posso **ver** as asas bem ali?” As palavras críticas aqui são “realmente” e “ver”.

199 — Explicar vs. justificar

O poema “Lamia” (1819) de John Keats [1] certamente merece algum tipo de prêmio de Poesia Mais Famosamente Irritante:

> Não fogem todos os encantos
>
> Ao mero toque da fria filosofia?
>
> Havia um arco-íris terrível outrora no céu:
>
> Conhecemos sua trama, sua textura; ela é dada
>
> No catálogo enfadonho das coisas comuns.
>
> A filosofia cortará as asas de um Anjo,
>
> Conquistará todos os mistérios por regra e linha,
>
> Esvaziará o ar assombrado e a mina dos gnomos—
>
> Destecerá um arco-íris[19]…

Minha resposta habitual termina com a frase: “Se não pudermos aprender a ter alegria no meramente real, nossas vidas serão de fato vazias”. Elaborarei sobre isso mais tarde.

Aqui tenho um ponto diferente em mente. Vamos apenas considerar os versos:

> Esvaziará o ar assombrado e a mina dos gnomos—
>
> *Destecerá um arco-íris…*

Aparentemente, “o mero toque da fria filosofia”, ou seja, a verdade, destruiu:

- Assombrações no ar;
- Gnomos na mina;
- Arco-íris.

---

19 NT. Texto original em inglês.
. . . Do not all charms fly
At the mere touch of cold philosophy?
There was an awful rainbow once in heaven:
We know her woof, her texture; she is given
In the dull catalogue of common things.
Philosophy will clip an Angel’s wings,
Conquer all mysteries by rule and line,
Empty the haunted air, and gnomed mine—
Unweave a rainbow . . .

O que nos faz lembrar um trecho de verso bem diferente:

Uma dessas coisas

Não é como as outras

Uma dessas coisas

Não pertence.

O ar foi esvaziado de suas assombrações, e a mina foi desgnomizada — mas o arco-íris continua lá!

Em "Corrigindo uma Pergunta Errada", escrevi:

Traçando de volta a cadeia de causalidade, passo a passo, descubro que minha crença de que estou usando meias é completamente explicada pelo fato de que estou usando meias... Por outro lado, se vejo uma miragem de um lago no deserto, a explicação causal correta da minha visão não envolve o fato de nenhum lago real no deserto. Neste caso, minha crença no lago não é apenas explicada, mas completamente eliminada.

O arco-íris foi explicado. As assombrações no ar e os gnomos na mina foram completamente eliminados.

Acho que esta é a distinção fundamental que os anti-reducionistas não entendem sobre o reducionismo.

Você pode ver essa falha em compreender a distinção na objeção clássica ao reducionismo:

Se o reducionismo estiver correto, então até mesmo sua crença no reducionismo é apenas o mero resultado do movimento de moléculas — por que eu deveria ouvir qualquer coisa que você diz?

A palavra-chave, no texto acima, é "mero"; uma palavra que implica que aceitar o reducionismo eliminaria completamente todos os processos de raciocínio que levaram à minha aceitação do reducionismo, da mesma forma que uma ilusão de ótica é completamente eliminada.

Mas você pode explicar como um processo cognitivo funciona sem que ele seja "mero"! Minha crença de que estou usando meias é um mero resultado do meu córtex visual reconstruindo impulsos nervosos enviados da minha retina que recebeu fótons refletidos das minhas meias... o que quer dizer, segundo o reducionismo científico, que minha crença de que estou usando meias é um mero resultado do fato de que estou usando meias.

O que poderia estar acontecendo nas mentes dos anti-reducionistas, de tal forma que eles colocariam arco-íris e crença no reducionismo na mesma categoria que assombrações e gnomos?

Várias coisas estão acontecendo simultaneamente. Mas por enquanto, vamos nos concentrar na ideia básica introduzida em um ensaio anterior: A Falácia da Projeção Mental entre um mapa multinível e um território mononível.

(Ou seja: não há como modelar um Boeing 747 quark por quark, então você tem que usar um mapa multinível com representações cognitivas explícitas de asas, fluxo de ar e assim por diante. Isso não significa que haja um território multinível. As verdadeiras leis da física, até onde sabemos, são apenas sobre campos de partículas elementares.)

Acho que quando os físicos dizem "Não existem arco-íris fundamentais", os anti-reducionistas ouvem "Não existem arco-íris".

Se você não consegue diferenciar o mapa multinível do território mononível, então quando alguém tenta explicar que o arco-íris não é um fenômeno físico fundamental, a aceitação disso será como se apagar o arco-íris do seu mapa multinível, o que é como apagar o arco-íris do mundo.

Quando a Ciência diz “tigres não são partículas elementares, eles são feitos de quarks”, o anti-reducionista ouve isso como o mesmo tipo de descarte que “procuramos um dragão na sua garagem, mas só havia ar vazio”.

O que os cientistas fizeram com os arco-íris e o que os cientistas fizeram com os gnomos pareceu aparentemente o mesmo para Keats...

Em apoio a esta sub-tese, eu deliberadamente usei várias frases, na minha discussão do poema de Keats, que eram Falácias de Projeção Mental. Se você não notou, isso pareceria argumentar que tais falácias são comuns o suficiente para passar despercebidas.

Por exemplo:

O ar foi esvaziado de suas assombrações, e a mina foi desgnomizada — mas o arco-íris continua lá!

Na verdade, a Ciência esvaziou o modelo de ar da crença em assombrações, e esvaziou o mapa da mina das representações de gnomos. A Ciência não realmente — como o próprio poema de Keats sugeriria — pegou asas reais de Anjo e as destruiu com um toque frio de verdade. Na realidade, nunca houveram assombrações no ar ou gnomos na mina.

Outro exemplo:

O que os cientistas fizeram com os arco-íris e o que os cientistas fizeram com os gnomos pareceu aparentemente o mesmo para Keats.

Os cientistas não fizeram nada com os gnomos, apenas com os “gnomos”. A citação não é o referente.

Mas se você cometer a Falácia da Projeção Mental — e por padrão, nossas crenças simplesmente parecem ser a maneira como o mundo é — então no tempo T = 0, as minas (aparentemente) contêm gnomos; no tempo T = 1, um cientista dança pela cena, e no tempo T = 2, as minas (aparentemente) estão vazias. Claramente, costumava haver gnomos lá, mas o cientista os matou.

Cientista mau! Nada de poemas para você, matador de gnomos!

Bem, essa é a sensação, se você se apega emocionalmente aos gnomos, e então um cientista diz não haver gnomos. É preciso uma mente forte, uma honestidade profunda e um esforço deliberado para dizer, neste ponto, “Aquilo que pode ser destruído pela verdade deve ser”, e “O cientista não levou os gnomos embora, apenas levou minha ilusão embora”, e “Eu nunca tive um título justo à crença em gnomos em primeiro lugar; não fui privado de nada que eu legitimamente possuísse”, e “Se existem gnomos, desejo acreditar que existem gnomos; se não existem gnomos, desejo acreditar que não existem gnomos; que eu não me apegue a crenças que posso não querer”, e todas as outras coisas que os racionalistas devem dizer em tais ocasiões.

Mas com o arco-íris nem é necessário ir tão longe. O arco-íris continua lá!

Referências

[1] John Keats, “Lamia,” The Poetical Works of John Keats (London: Macmillan) (1884).

200 — Falso reducionismo

Havia um arco-íris terrível outrora no céu:

Conhecemos sua trama, sua textura; ela é dada

No catálogo enfadonho das coisas comuns.

—John Keats, Lamia

Estou supondo — embora seja apenas um palpite — que o próprio Keats não conhecia a trama e a textura do arco-íris. Não da maneira que Newton entendia os arco-íris. Talvez nem um pouco. Pode ser que Keats tenha apenas lido, em algum lugar, que Newton havia explicado o arco-íris como "luz refletida das gotas de chuva" —

— o que já era conhecido no século XIII. Newton apenas acrescentou um refinamento ao mostrar que a luz era decomposta em partes coloridas, em vez de transformada em cor. Mas isso colocou os arco-íris de volta nas manchetes. E assim Keats, com Charles Lamb, William Wordsworth e Benjamin Haydon, brindou à "confusão à memória de Newton" porque "ele destruiu a poesia do arco-íris ao reduzi-lo a um prisma". Essa é uma razão para suspeitar que Keats não entendia o assunto muito profundamente.

Estou supondo, embora seja apenas um palpite, que Keats não poderia ter esboçado no papel por que os arco-íris só aparecem quando o Sol está atrás de sua cabeça, ou por que o arco-íris é um arco de círculo.

Se for assim, Keats tinha uma Explicação Falsa. Neste caso, uma redução falsa. Ele havia sido informado de que o arco-íris tinha sido reduzido, mas, na verdade, não havia sido reduzido em seu modelo de mundo.

Esta é outra daquelas distinções que os anti-reducionistas não conseguem entender — a diferença entre professar o fato simples de que algo é redutível e realmente vê-lo.

Nisto, os anti-reducionistas não são muito culpados, ao fazer parte de um problema geral.

Já discorri sobre o saber aparente que não é saber, sobre as crenças que não se referem aos seus supostos objetos, mas apenas as gravações para serem recitadas em sala de aula, as palavras que atuam como placas de pare para a curiosidade em vez de respostas, e a tecnobaboseira que apenas transmite pertencimento ao gênero literário da "ciência"…

Há uma grande distinção entre ser capaz de ver de onde vem o arco-íris, e brincar com prismas para confirmá-lo e talvez fazer um arco-íris você mesmo borrifando gotas de água —

— versus algum filósofo de cara fechada simplesmente lhe dizer: "Não, não há nada de especial no arco-íris. Você não ouviu? Os cientistas já o explicaram. Algo a ver com gotas de chuva, ou seja, lá o que for. Nada para se empolgar."

Acho que essa distinção provavelmente explica boa parte do vazio existencial mortal que supostamente acompanha o reducionismo científico.

Você tem que interpretar a experiência dos anti-reducionistas com o “reducionismo”, não em termos de eles realmente verem como os arco-íris funcionam, não em termos de eles terem o “Aha!” crítico, mas em termos de eles serem ditos que a senha é “Ciência”. O efeito é apenas mover os arco-íris para um gênero literário diferente — um gênero literário que eles foram ensinados a considerar chato.

Para eles, o efeito de ouvir “A Ciência explicou os arco-íris!” é pendurar uma placa sobre os arco-íris dizendo: “este fenômeno foi rotulado como chato por ordem do Conselho de Críticos Literários Sofisticados. Siga em frente.”

E é só isso que a placa diz: apenas isso, e nada mais.

Assim, os críticos literários têm seus gnomos arrancados à força; não dissolvidos em compreensão, mas removidos por ordem categórica da autoridade. Não lhes é dada nenhuma beleza para substituir o ar desassombrado, nenhuma compreensão genuína que poderia ser interessante por si só. Apenas um rótulo dizendo: “Ha! Você achou que os arco-íris eram bonitos? Seu tolo, não sofisticado. Isso faz parte do gênero literário da ciência, de palavras secas, solenes e incompreensíveis.”

É assim que os anti-reducionistas experimentam o “reducionismo”.

Bem, não podemos culpar Keats, o pobre rapaz provavelmente não foi criado corretamente.

Mas ele ousou brindar à “Confusão à memória de Newton”?

Proponho “À memória da confusão de Keats” como um brinde para os racionalistas.

Saúde.

201 — Poetas da savana

> Poetas dizem que a ciência rouba a beleza das estrelas - meros glóbulos de átomos gasosos. Nada é "mero". Eu também posso ver as estrelas em uma noite no deserto, e senti-las. Mas eu vejo menos ou mais?
>
> A vastidão dos céus expande minha imaginação - preso neste carrossel, meu pequeno olho pode captar a luz de um milhão de anos. Um vasto padrão - do qual faço parte - talvez meu material tenha sido expelido de alguma estrela esquecida, como uma está expelindo agora. Ou vê-las com o maior olho do Palomar, todas se afastando de um ponto de partida comum, quando talvez estivessem todas juntas. Qual é o padrão, ou o significado, ou o porquê? Não prejudica o mistério saber um pouco sobre ele.
>
> Pois a verdade é muito mais maravilhosa do que qualquer artista do passado imaginou! Por que os poetas do presente não falam disso?
>
> Que homens são os poetas que podem falar de Júpiter se ele fosse como um homem, mas se ele é uma imensa esfera giratória de metano e amônia devem ficar em silêncio?[20]
>
> — Richard Feynman, The Feynman Lectures on Physics, [1]
>
> Vol I, p. 3–6 (quebras de linha adicionadas)

Essa é uma pergunta real, ali na última linha: que tipo de poeta pode escrever sobre Júpiter, o deus, mas não sobre Júpiter, a imensa esfera? Quer Feynman tenha ou não feito a pergunta retoricamente, ela tem uma resposta real:

> Se Júpiter fosse como nós, ele poderia se apaixonar, perder o amor e recuperar o amor.
>
> Se Júpiter fosse como nós, ele poderia lutar, ascender e ser derrubado.
>
> Se Júpiter fosse como nós, ele poderia rir, chorar ou dançar.

Se Júpiter é uma imensa esfera giratória de metano e amônia, é mais difícil para o poeta nos fazer sentir.

Há poetas e contadores de histórias que dizem que as Grandes Histórias são atemporais e nunca mudam, apenas são recontadas. Eles dizem, com orgulho, que Shakespeare e Sófocles estão ligados por laços de ofício mais fortes do que meros séculos; que os dois dramaturgos poderiam ter trocado de época sem sobressaltos.

Donald Brown certa vez compilou uma lista de mais de duzentos "universais humanos", encontrados

20 NT. Texto original em inglês. *Poets say science takes away from the beauty of the stars—mere globs of gas atoms. Nothing is "mere." I too can see the stars on a desert night, and feel them. But do I see less or more? The vastness of the heavens stretches my imagination—stuck on this carousel my little eye can catch one-million-year-old light. A vast pattern—of which I am a part—perhaps my stuff was belched from some forgotten star, as one is belching there. Or see them with the greater eye of Palomar, rushing all apart from some common starting point when they were perhaps all together. What is the pattern, or the meaning, or the why? It does not do harm to the mystery to know a little about it. For far more marvelous is the truth than any artists of the past imagined! Why do the poets of the present not speak of it? What men are poets who can speak of Jupiter if he were like a man, but if he is an immense spinning sphere of methane and ammonia must be silent?*

em todas (ou na grande maioria) das culturas humanas estudadas, de São Francisco aos !Kung[21] do deserto do Kalahari. O casamento está na lista, assim como a prevenção do incesto, o amor materno, a rivalidade entre irmãos, a música, a inveja, a dança, a narrativa, a estética e a magia ritual para curar os doentes, e a poesia em versos falados separados por pausas.

Ninguém que saiba algo sobre psicologia evolutiva poderia negar: as emoções mais fortes que temos estão profundamente gravadas em nosso sangue e ossos, cérebro e DNA.

Talvez fosse preciso alguns ajustes, mas você provavelmente poderia contar "Hamlet" sentado em volta de uma fogueira na savana ancestral.

Então, dá para entender por que John "Desteça um arco-íris" Keats poderia sentir que algo se perdeu ao ser informado de que o arco-íris era a luz do sol dispersa pelas gotas de chuva. Gotas de chuva não dançam.

No Antigo Testamento, há uma história sobre um dilúvio enviado por Deus que cobriu o mundo inteiro e levou à morte de todos os homens e mulheres terrivelmente culpados do mundo, juntamente com seus bebês horrivelmente culpados, mas Noé a construiu uma grande arca de madeira, e assim por diante, e depois a maioria da humanidade foi exterminada, Deus colocou arco-íris no céu como um sinal de que não faria isso novamente. Pelo menos não com água.

Você pode ver como Keats ficaria chocado que essa bela história fosse contrariada pela ciência moderna. Especialmente se (como descrevi no ensaio anterior) Keats não tivesse um entendimento real dos arco-íris, nenhum insight "Aha!" que pudesse ser fascinante por si só, para substituir o drama subtraído.

Ah, mas talvez Keats tivesse razão em ficar desapontado, mesmo que soubesse a matemática. A história bíblica do arco-íris é um conto de assassinato sanguinário e insanidade sorridente. Como algo sobre gotas de chuva e refração poderia substituir isso adequadamente? Gotas de chuva não gritam quando morrem.

Então a ciência tira o romance (diz o poeta romântico), e o que você recebe de volta nunca corresponde ao drama do original -

(isto é, a ilusão original)

-mesmo que você saiba as equações, porque as equações não são sobre emoções fortes.

Essa é a réplica mais forte que posso imaginar que qualquer poeta romântico poderia ter dito a Feynman, embora eu não me lembre de tê-la ouvido.

Você pode imaginar que eu não concorde com os poetas românticos. Então, minha própria postura é esta:

Não é necessário que Júpiter seja como um humano, porque os humanos são como humanos. Se Júpiter é uma imensa esfera giratória de metano e amônia, isso não significa que o amor e o ódio sejam esvaziados do universo. Ainda existem mentes amorosas e odiosas no universo. Nós.

Com mais de seis bilhões de nós na última contagem, Júpiter realmente precisa estar na lista de protagonistas em potencial?

Não é necessário contar as Grandes Histórias sobre planetas ou arco-íris. Elas se desenrolam em todo o nosso mundo, todos os dias. Todos os dias, alguém mata por vingança; todos os dias, alguém mata um amigo por engano; todos os dias, mais de cem mil pessoas se apaixonam. E mesmo que não fosse assim, você poderia escrever ficção sobre humanos - não sobre Júpiter.

---

21 NT. Os **!Kung** (também chamados Ju/'hoansi) são um povo indígena que vive sobretudo no Deserto de Kalahari, no sul da África. Pertencentes ao grupo San, são conhecidos por sua tradição caçadora-coletora e uso de línguas caracterizadas por sons de clique. Historicamente, mantêm forte conexão com o ambiente, praticando técnicas sofisticadas de sobrevivência em regiões áridas.

A Terra é antiga e já encenou as mesmas histórias muitas vezes sob o Sol. Eu me pergunto se não seria hora de algumas das Grandes Histórias mudarem. Para mim, pelo menos, a história chamada "Adeus" perdeu o seu charme.

As Grandes Histórias não são atemporais, porque a espécie humana não é atemporal. Volte longe o suficiente na evolução dos hominídeos e ninguém entenderá Hamlet. Volte longe o suficiente no tempo e você não encontrará nenhum cérebro.

As Grandes Histórias não são eternas, porque a espécie humana, *Homo sapiens sapiens*, não é eterna. Duvido sinceramente que tenhamos mais mil anos pela frente em nossa forma atual. Não digo isso com tristeza: acho que podemos fazer melhor.

Eu não gostaria de ver todas as Grandes Histórias perdidas completamente, em nosso futuro. Vejo muito pouca diferença entre esse resultado e o Sol caindo em um buraco negro.

Mas as Grandes Histórias em suas formas atuais já foram contadas, repetidas vezes. Não acho ruim se algumas delas mudarem de forma ou diversificarem seus finais.

"E viveram felizes para sempre" parece valer a pena tentar pelo menos uma vez.

As Grandes Histórias podem e devem se diversificar à medida que a humanidade cresce. Parte dessa ética é a ideia de que, quando encontramos estranheza, devemos respeitá-la o suficiente para contar sua história com verdade. Mesmo que isso torne a escrita da poesia um pouco mais difícil.

Se você é um poeta bom o suficiente para escrever uma ode a uma imensa esfera giratória de metano e amônia, você está escrevendo algo original, sobre uma parte recém-descoberta do universo real. Pode não ser tão dramático ou tão emocionante quanto Hamlet. Mas a história de Hamlet já foi contada! Se você escreve sobre Júpiter como se fosse um humano, está tornando nosso mapa do universo um pouco mais empobrecido em complexidade; você está forçando Júpiter no molde de todas as histórias que já foram contadas sobre a Terra.

O poema de James Thomson, "Um Poema Sagrado à Memória de Sir Isaac Newton", que elogia o arco-íris pelo que ele realmente é - você pode argumentar se o poema de Thomson é ou não tão emocionante quanto a Lâmia de John Keats, que foi amada e perdida. Mas contos de amor, perda e cinismo já haviam sido contados, muito longe na Grécia antiga, e sem dúvida muitas vezes antes. Até que entendêssemos o arco-íris como algo diferente dos contos de magia em forma humana, a verdadeira história do arco-íris não poderia ser poetizada.

A fronteira entre a ficção científica e a ópera espacial já foi traçada da seguinte forma: se você pode pegar o enredo de uma história e colocá-lo de volta no Velho Oeste ou na Idade Média, sem o alterar, então não é ficção científica real. Na verdadeira ficção científica, a ciência é intrinsecamente parte do enredo; você não pode mover a história do espaço para a savana, não sem perder alguma coisa.

Richard Feynman perguntou: "Que homens são os poetas que podem falar de Júpiter se ele fosse como um homem, mas se ele é uma imensa esfera giratória de metano e amônia devem ficar em silêncio?"

Eles são poetas da savana, que só podem contar histórias que fariam sentido em volta de uma fogueira há dez mil anos. Poetas da savana, que só podem contar as Grandes Histórias em suas formas clássicas, e nada mais.

Referências

[1] Richard P. Feynman, Robert B. Leighton, and Matthew L. Sands, The Feynman Lectures on Physics, 3 vols. (Reading, MA: Addison-Wesley, 1963).

Parte Q — Alegria no meramente real

202 — Alegria no meramente real

…Não fogem todos os encantos

Ao mero toque da fria filosofia?

Havia um arco-íris terrível outrora no céu:

Conhecemos sua trama, sua textura; ela é dada

No catálogo enfadonho das coisas comuns.

— John Keats, Lâmia

Nada é “mero”.

— Richard Feynman

É preciso admirar essa frase, “catálogo enfadonho das coisas comuns”. O que exatamente entra nesse catálogo? Além de arco-íris, é claro?

Ora, coisas que são mundanas, naturalmente. Coisas que são normais; coisas que não são mágicas; coisas conhecidas ou cognoscíveis; coisas que seguem as regras (ou que seguem quaisquer regras, o que as torna entediantes); coisas que fazem parte do universo ordinário; coisas que são, em uma palavra, reais.

Isso, sim, é o que chamo de se preparar para uma queda.

Nesse ritmo, mais cedo ou mais tarde você vai se decepcionar com tudo — ou vai descobrir que não existe, ou pior ainda, descobrirá ser real.

Se não pudermos sentir alegria nas coisas que são meramente reais, nossas vidas sempre estarão vazias.

Por qual pecado os arco-íris são rebaixados ao catálogo enfadonho das coisas comuns? Pelo pecado de terem uma explicação científica. “Conhecemos sua trama, sua textura”, diz Keats — um uso interessante da palavra “conhecemos”, porque suspeito que o próprio Keats não conhecia a explicação. Suspeito que apenas ser informado que outra pessoa sabia já era demais para ele. Suspeito que a mera noção de que arco-íris fossem cientificamente explicáveis em princípio já seria demais. E se Keats não pensava assim, bem, conheço muitas pessoas que pensam.

Já observei antes que nada é inerentemente misterioso — nada que realmente exista, isto é. Se sou ignorante sobre um fenômeno, isso é um fato sobre meu estado mental, não um fato sobre o fenômeno; adorar um fenômeno porque ele parece tão maravilhosamente misterioso é adorar sua própria ignorância; um mapa em branco não corresponde a um território em branco, é apenas um lugar que ainda não visitamos, etc., etc. …

O que significa dizer que tudo — tudo que realmente existe — está sujeito a acabar no “catálogo enfadonho das coisas comuns”, mais cedo ou mais tarde.

Sua escolha é:

- Decidir que as coisas podem ser não mágicas, cognoscíveis, cientificamente explicáveis — em uma palavra, reais — e ainda assim valer a pena se importar com elas;
- Ou passar o resto da vida sofrendo de um tédio existencial irresolúvel.

(O autoengano pode ser uma opção para outros, mas não para você.)

Isso coloca uma perspectiva bem diferente sobre o hábito bizarro praticado por aquelas pessoas estranhas chamadas cientistas, em que elas de repente se fascinam por fiapos de bolso ou excrementos de pássaros, ou arco-íris, ou alguma outra coisa comum que pessoas sofisticadas e entediadas com o mundo nunca dariam uma segunda olhada.

Pode-se dizer que os cientistas — pelo menos alguns cientistas — são aquelas pessoas que são, em princípio, capazes de aproveitar a vida no universo real.

203 — Alegria na descoberta

> Newton foi o maior gênio que já viveu, e o mais sortudo; pois não podemos encontrar mais de uma vez um sistema do mundo para estabelecer[22].
>
> —Lagrange

Eu me divirto mais descobrindo as coisas por mim mesmo do que lendo sobre elas em livros didáticos. Isso é certo e adequado, e apenas o esperado.

Mas descobrir algo que ninguém mais sabe - ser o primeiro a desvendar o segredo -

Há uma história de que um dos primeiros homens a perceber que as estrelas queimavam por fusão - atribuições plausíveis que vi são para Fritz Houtermans e Hans Bethe - estava saindo com sua namorada de uma noite, e ela comentou sobre como as estrelas eram bonitas, e ele respondeu: “Sim, e nesse momento, sou o único homem no mundo que sabe por que elas brilham.”

É atestado por numerosas fontes que essa experiência, de ser a primeira pessoa a resolver um grande mistério, é uma onda tremenda. É provavelmente a experiência mais próxima que você pode ter de usar drogas, sem usar drogas - embora eu não saiba.

Isso não pode ser saudável.

Não que eu esteja me opondo à euforia. É a cláusula de exclusividade que me incomoda. Por que uma descoberta deveria valer menos, apenas porque alguém já sabe a resposta?

A interpretação mais caridosa que posso dar à psicologia é que você não luta com um único problema por meses ou anos se puder resolvê-lo facilmente consultando um livro. E que aquela onda tremenda surge quando você examina o problema de todos os ângulos possíveis e ainda não consegue encontrar uma solução; e então, você o analisa novamente, usando todas as ideias e evidências disponíveis - avançando gradualmente - até que, finalmente, quando desvenda o enigma, todas as peças soltas e questões sem resposta se encaixam perfeitamente, como resolver uma dúzia de mistérios de assassinato em um quarto fechado com uma única pista.

E mais, a compreensão que você obtém é uma compreensão real - compreensão que abrange todas as pistas que você estudou para resolver o problema, quando você ainda não sabia a resposta. Compreensão que vem de fazer perguntas dia após dia e se preocupar com elas; compreensão que ninguém mais pode obter (não importa o quanto você diga a eles a resposta) a menos que passem meses estudando o problema em seu contexto histórico, mesmo após ter sido resolvido - e mesmo assim, eles não terão a onda de resolvê-lo de uma só vez.

Essa é uma possível razão pela qual James Clerk Maxwell pode ter se divertido mais descobrindo as equações de Maxwell do que você se divertiu lendo sobre elas.

Uma leitura um pouco menos caridosa é que a onda tremenda vem do que é chamado, na polidez da

22 NT. Texto original em inglês. *Newton was the greatest genius who ever lived, and the most fortunate; for we cannot find more than once a system of the world to establish.*

psicologia social, de "comprometimento" e "consistência" e "dissonância cognitiva"; a parte em que valorizamos algo mais, apenas porque foi preciso mais trabalho para obtê-lo. Os estudos que mostram que submeter os calouros da fraternidade a uma iniciação mais severa, faz com que eles fiquem mais convencidos do valor da fraternidade - vinho idêntico em garrafas de preço mais alto sendo classificado como tendo um sabor melhor - esse tipo de coisa.

Claro, se você se diverte mais resolvendo um quebra-cabeça do que ser informado de sua resposta, porque você gosta de fazer o trabalho cognitivo por si só, não há nada de errado com isso. A leitura menos caridosa seria se cobrar R$ 100 para ser informado da resposta a um quebra-cabeça fizesse você pensar que a resposta era mais interessante, valiosa, importante, surpreendente, etc., do que se você obtivesse a resposta de graça.

(Eu suspeito fortemente que uma grande parte do problema de relações-públicas da ciência na população em geral são pessoas que instintivamente acreditam que se o conhecimento for dado, ele não pode ser importante. Se você tivesse que passar por um ritual de iniciação temível para ser informado da verdade sobre a evolução, talvez as pessoas ficassem mais satisfeitas com a resposta.)

A leitura realmente não caridosa é que a alegria da primeira descoberta é pelo status. Competição. Escassez. Vencer todos os outros. Não importa se você tem uma casa de três quartos ou uma casa de quatro quartos, o que importa é ter uma casa maior do que a dos vizinhos. Uma casa de dois quartos seria suficiente, se você pudesse garantir que os vizinhos tivessem ainda menos.

Eu não me oponho à competição como uma questão de princípio. Eu não acho que o jogo de Go seja bárbaro e deva ser suprimido, mesmo sendo de soma zero. Mas se a alegria eufórica da descoberta científica tiver que ser sobre escassez, isso significa que ela está disponível apenas para uma pessoa por civilização para qualquer verdade dada.

Se a alegria da descoberta científica for única por descoberta, então, de uma perspectiva teórica da diversão, Newton provavelmente usou um incremento substancial do total de Diversão da Física disponível ao longo de toda a história da vida inteligente originada na Terra. Aquele egoísta explicou as órbitas dos planetas e as marés.

E realmente a situação é ainda pior do que isso, porque no Modelo Padrão da física (descoberto por egoístas que estragaram o quebra-cabeça para todos os outros) o universo é espacialmente infinito, inflacionariamente ramificado e ramificado por decoerência, os quais são pelo menos três maneiras diferentes de a Realidade ser exponencial ou infinitamente grande.

Então, alienígenas, ou Newtons alternativos, ou apenas duplicatas de Tegmark de Newton, todos podem ter descoberto a gravidade antes do nosso Newton - se você acredita que "antes" significa algo em relação a esses tipos de separações.

Quando esse pensamento me ocorreu pela primeira vez, eu realmente o achei bastante edificante. Uma vez que percebi que alguém, em algum lugar nas extensões do espaço e do tempo, já sabe a resposta para qualquer pergunta respondível - até mesmo perguntas sobre biologia e história; existem outras Terras decoerentes - então percebi como era bobo pensar como se a alegria da descoberta devesse ser limitada a uma pessoa.

Isso se torna uma fonte totalmente inevitável de angústia existencial insolúvel, e eu considero isso um *reductio ad absurdum*[23].

A solução consistente que mantém a possibilidade de diversão é parar de se preocupar com o que as outras pessoas sabem. Se você não sabe a resposta, é um mistério para você. Se você pode levantar a mão e cerrar os dedos em um punho, e não tem ideia de como seu cérebro está fazendo isso - ou mesmo quais

23 NT. Latim. ***Reductio ad absurdum*** (ou redução ao absurdo) é um método de demonstração lógica que consiste em provar a veracidade de uma proposição ao mostrar que a negação dela leva a uma contradição ou impossibilidade. Desse modo, conclui-se que a proposição original deve ser verdadeira.

músculos exatos estão sob sua pele - você tem que se considerar tão ignorante quanto um caçador-coletor. Claro, alguém sabe a resposta - mas nos dias de caçador-coletor, alguém em uma Terra alternativa, ou, aliás, alguém no futuro, sabia qual era a resposta. O mistério, e a alegria de descobrir, é algo pessoal, ou nem existem - e eu prefiro dizer que é pessoal.

A alegria de ajudar sua civilização dizendo a ela algo que ela ainda não sabe tende a ser única por descoberta por civilização; esse tipo de valor é conservado, assim como os Prêmios Nobel. E a perspectiva dessa recompensa pode ser o que é preciso para mantê-lo focado em um problema pelos anos necessários para desenvolver um entendimento realmente profundo; além disso, trabalhar em um problema desconhecido para sua civilização é uma maneira infalível de evitar ler spoilers.

Mas como parte de meu projeto geral de desmitificar a ideia de que os racionalistas não se divertem, quero restaurar o encanto e o mistério a cada aspecto do mundo que você pessoalmente não compreende, independentemente de qualquer outro conhecimento que possa existir, distante no espaço e no tempo, ou mesmo na mente de seu vizinho. Se você não sabe, é um mistério. E agora pense em quantas coisas você não sabe! (Se você não consegue pensar em nada, você tem outros problemas.) O mundo não é de repente um lugar muito mais misterioso, mágico e interessante? Como se você tivesse sido transportado para uma dimensão alternativa e tivesse que aprender todas as regras do zero?

> Um amigo me disse uma vez que olho para o mundo como se nunca o tivesse visto antes. Eu pensei, que elogio legal... Espere! Eu nunca vi isso antes! O quê - todo mundo já o viu antes[24]?
>
> — Ran Prieur

24 NT. Texto original em inglês. A friend once told me that I look at the world as if I've never seen it before. I thought, that's a nice compliment . . . Wait! I never have seen it before! What—did everyone else get a preview?

204 — Prenda-se à realidade

Então, talvez você esteja lendo tudo isso e se perguntando: “Sim, mas o que isso tem a ver com reducionismo?”

Em parte, trata-se de deixar uma linha de recuo. Não é fácil desmontar algo importante em componentes quando você está convencido de que isso remove a magia do mundo, desfaz o arco-íris. Pretendo desmontar certas coisas neste livro; e prefiro não criar uma angústia existencial desnecessária.

Em parte, é a cruzada contra a Racionalidade de Hollywood, o conceito de que entender o arco-íris subtrai sua beleza. O arco-íris ainda é bonito, e você ganha a beleza da física.

Mas, ainda mais profundamente, é uma dessas coisas sutis do núcleo oculto da racionalidade. Você sabe, o tipo de coisa em que começo a falar sobre “o Caminho”. É sobre se prender à realidade.

Em um dos livros da série Duna de Frank Herbert, se bem me lembro, é dito que um Revelador da Verdade ganha a capacidade de detectar mentiras nos outros ao falar sempre a verdade, de modo que forma uma relação com a verdade cuja violação ele pode sentir. Não funcionaria, mas ainda acho que é um dos pensamentos mais bonitos da ficção. Pelo menos, para chegar perto da verdade, você precisa estar disposto a se pressionar contra a realidade o mais firmemente possível, sem recuar ou se rebaixar.

Você pode ver o tema de se prender à realidade em “Loterias: Um Desperdício de Esperança”. Entender que bilhetes de loteria têm utilidade esperada negativa não significa que você desiste da esperança de ser rico. Significa que você para de desperdiçar essa esperança em bilhetes de loteria. Você coloca a esperança em seu trabalho, sua escola, sua startup, seu negócio no eBay; e se você realmente não tem nada que valha a pena esperar, então talvez seja hora de começar a procurar.

Não é contra os sonhos que eu me oponho, apenas contra os sonhos impossíveis. A loteria não é impossível, mas é uma quase-impossibilidade não acionável. Não é que ganhar na loteria seja extremamente difícil—requer um esforço desesperado—mas que o trabalho não é o problema.

Digo tudo isso para exemplificar a ideia de pegar a energia emocional que está fluindo para lugar nenhum e vinculá-la aos domínios da realidade.

Isso não significa estabelecer metas que sejam baixas o suficiente para serem “realistas”, ou seja, fáceis e seguras e aprovadas pelos pais. Talvez isso seja um bom conselho no seu caso pessoal, não sei, mas não sou eu quem deve dizer isso.

O que quero dizer é que você pode investir energia emocional em arco-íris, mesmo que eles se revelem não serem mágicos. O futuro é sempre absurdo, mas nunca é irreal.

O estereótipo da Racionalidade de Hollywood é que “racional = sem emoção”; quanto mais razoável você é, mais de suas emoções a Razão inevitavelmente destrói. Em “Sentindo-se Racional” eu contrasto isso com “Aquilo que pode ser destruído pela verdade deve ser” e “Aquilo que a verdade nutre deve prosperar.” Quando você chega à sua melhor visão da verdade, não há nada de irracional nas emoções que você sente como resultado disso — as emoções não podem ser destruídas pela verdade, portanto elas não devem ser irracionais.

Então, em vez de destruir energias emocionais associadas a más explicações para arco-íris, como o estereótipo da Racionalidade de Hollywood diria, vamos redirecionar essas energias emocionais para a realidade—prendê-las a crenças que são tão verdadeiras quanto podemos torná-las.

Quer voar? Não desista de voar. Desista de poções voadoras e construa você mesmo um avião.

Lembre-se do tema de "Pense como a Realidade", onde falei sobre como e quando a física parece contraintuitiva, você tem que aceitar que não é a física que é estranha, é você?

O que estou falando agora é algo assim, só que com emoções em vez de hipóteses—vincular seus sentimentos ao mundo real. Não ao mundo "realista" do dia a dia. Eu seria um hipócrita uivante se dissesse para você calar a boca e fazer seu dever de casa. Quero dizer, o mundo real de verdade, o universo legal, que inclui absurdos como pousos na Lua e a evolução da inteligência humana. Só não há nenhuma magia, em lugar nenhum, jamais.

É um meme da Racionalidade de Hollywood que "a Ciência tira a diversão da vida."

A Ciência coloca a diversão de volta na vida.

A racionalidade direciona suas energias emocionais para o universo, em vez de para outro lugar.

205 — Se você exige magia, a magia não vai ajudar

> A maioria das bruxas não acredita em deuses. Elas sabem que os deuses existem, é claro. Até lidam com eles ocasionalmente. Mas não acreditam neles. Elas os conhecem bem demais. Seria como acreditar no carteiro[25].
>
> — Terry Pratchett, *Witches Abroad* (Feiticeiras no exterior) [1]

Era uma vez, eu estava ponderando sobre a filosofia das histórias de fantasia—

E antes que alguém me repreenda por minha "falha em entender do que trata a fantasia", deixe-me dizer o seguinte: fui criado em uma casa de ficção científica e fantasia. Leio histórias de fantasia desde os cinco anos. Ocasionalmente, tento escrever histórias de fantasia. E não sou o tipo de pessoa que tenta escrever para um gênero sem ponderar sua filosofia. De onde você acha que vêm as ideias para histórias?

De qualquer forma:

Eu estava refletindo sobre a filosofia das histórias de fantasia, e ocorreu-me que se realmente houvesse dragões em nosso mundo — se você pudesse ir ao zoológico, ou até mesmo a uma montanha distante, e encontrar um dragão que cospe fogo — enquanto ninguém jamais tivesse visto uma zebra, então nossas histórias de fantasia estariam repletas de zebras, enquanto dragões seriam desinteressantes.

Isso, sim, é o que chamo de se encurralar, não é? A grama é sempre mais verde do outro lado da irrealidade.

Em um dos enredos padrão de fantasia, um protagonista de nossa Terra, um personagem simpático com notas ruins ou uma hipoteca esmagadora, mas ainda com um bom coração, de repente se encontra em um mundo onde a magia opera no lugar da ciência. O protagonista passa frequentemente a praticar magia e se torna, com o tempo, um feiticeiro (superpoderoso).

Agora, eis a questão — e sim, é um pouco cruel, mas acho que precisa ser feita: presumivelmente, a maioria dos leitores desses romances se vê no lugar do protagonista, fantasiando sobre sua própria aquisição de feitiçaria. Desejando magia. E, excluindo demografias improváveis, a maioria dos leitores desses romances não são cientistas.

Nascidos em um mundo de ciência, eles não se tornaram cientistas. O que os faz pensar que, em um mundo de magia, agiriam de forma diferente?

Se eles não têm a atitude científica, de que nada é "mero" — a capacidade de se interessar por coisas meramente reais — como a magia os ajudará? Se eles realmente tivessem magia, ela seria meramente real e perderia o encanto da inatingibilidade. Eles poderiam ficar empolgados no início, mas (como os ganhadores da loteria que, seis meses depois, não estão nem de longe tão felizes quanto esperavam estar), a empolgação logo passaria. Provavelmente assim que tinham que estudar de fato os feitiços.

A menos que possam encontrar a capacidade de sentir alegria em coisas que são meramente reais. De ficar tão empolgados com voo livre quanto com montar um dragão; de ficar tão empolgados em fazer luz

25 NT. Texto original em inglês. *Most witches don't believe in gods. They know that the gods exist, of course. They even deal with them occasionally. But they don't believe in them. They know them too well. It would be like believing in the postman.*

com eletricidade quanto em fazer luz com magia… mesmo que isso exija um pouco de estudo…

Não me entenda mal, não estou menosprezando os dragões. Quem sabe, talvez até criemos alguns, um dia desses.

Mas se você não pode aproveitar o voo livre, mesmo que ele seja meramente real, então assim que os dragões se tornarem reais, você não ficará mais empolgado com dragões do que está com voo livre.

Você acha que preferiria viver no Futuro a viver no presente? Essa é uma preferência bastante compreensível. As coisas parecem estar melhorando com o tempo.

Mas não se esqueça de que este é o Futuro, em relação à Idade das Trevas de mil anos atrás. Você tem oportunidades inimagináveis até mesmo para reis.

Se a tendência continuar, o Futuro pode ser um lugar ótimo para se viver. Mas se você chegar ao Futuro, o que encontrará quando chegar lá será outro Agora. Se você não tem a capacidade básica de aproveitar estar em um Agora — se sua energia emocional só pode ir para o Futuro, se você só pode esperar por um amanhã melhor — então nenhuma quantidade de tempo passando pode ajudá-lo.

(Sim, no Futuro poderia haver uma pílula que resolvesse o problema emocional de sempre olhar para o Futuro. Não acho que isso invalide meu ponto básico, o qual é sobre que tipo de pílulas deveríamos querer tomar.)

Matthew C., comentando no Less Wrong, parece muito empolgado com uma "teoria" informalmente especificada por Rupert Sheldrake que "explica" fenômenos que não demandam explicação, como o dobramento de proteínas e a simetria dos flocos de neve. Mas por que Matthew C. não está tão empolgado com, digamos, a Relatividade Especial? A Relatividade Especial é realmente conhecida por ser uma lei, então por que não é ainda mais empolgante? A vantagem de ficar empolgado com uma lei que já se sabe ser verdadeira é que você sabe que sua empolgação não será desperdiçada.

Se a teoria de Sheldrake fosse uma verdade aceita ensinada nas escolas primárias, Matthew C. não se importaria com ela. Ou por que mais Matthew C. estaria fascinado por essa única lei particular que ele acredita ser uma lei da física, mais do que todas as outras leis?

A pior catástrofe que você poderia infligir à comunidade da Nova Era seria fazer com que seus rituais começassem a funcionar de forma confiável, e que Óvnis realmente aparecessem nos céus. Qual seria o sentido de acreditar em alienígenas se eles estivessem simplesmente lá, e todos os outros também pudessem vê-los? Em um mundo onde poderes psíquicos fossem meramente reais, os adeptos da Nova Era não acreditariam em poderes psíquicos, assim como ninguém se importa o suficiente com a gravidade para acreditar nela. (Exceto os cientistas, é claro.)

Por que sou tão negativo em relação à magia? Seria errado a magia existir?

Na verdade, não sou negativo em relação à magia. Lembre-se, eu ocasionalmente tento escrever histórias de fantasia. Mas estou incomodado com essa psicologia que, se nascesse em um mundo onde feitiços e poções realmente funcionassem, ansiaria por um mundo onde bens domésticos fossem produzidos abundantemente por linhas de montagem.

Parte de se vincular à realidade, em um nível emocional e intelectual, é aceitar que você realmente vive aqui. Só então você pode ver isto, seu mundo, e quaisquer oportunidades que ele ofereça para você, sem desejar que sua visão se dissipe.

Para não ser muito sutil, não encontrei falta de dragões para lutar, ou magias para dominar, neste mundo em que nasci. Se eu fosse transportado para um desses romances de fantasia, não ficaria surpreso em me encontrar estudando a feitiçaria suprema proibida—

— Por que ser transportado para um mundo mágico mudaria alguma coisa?

Não é onde você está, é quem você é.

Então, lembre-se da Litania Contra Ser Transportado Para Um Universo Alternativo:

Se serei feliz em algum lugar,

Ou alcançarei grandeza em algum lugar,

Ou aprenderei verdadeiros segredos em algum lugar,

Ou salvarei o mundo em algum lugar,

Ou sentirei fortemente algum lugar,

Ou ajudarei pessoas em algum lugar…

É melhor fazer isso na realidade.

Referências

[1] Terry Pratchett, Witches Abroad (London: Corgi Books, 1992).

206 — Magia mundana

Acho que parte do *ethos* racionalista é se prender emocionalmente a um universo absolutamente legalista e reducionista—um universo que não contém coisas sobrenaturais como almas ou magia—e derramar toda a sua esperança e todo o seu cuidado nesse universo meramente real e em suas possibilidades, sem decepção.

Há um velho truque para combater o *dukkha* em que você faz uma lista de coisas pelas quais é grato, como um teto sobre sua cabeça.

Então, por que não fazer uma lista de habilidades que você tem que seriam incrivelmente legais se fossem mágicas, ou se apenas alguns escolhidos as tivessem?

Por exemplo, suponha que, em vez de um olho, você possuísse um segundo olho mágico embutido na testa. E esse segundo olho permitisse que você visse na terceira dimensão—de modo que você pudesse de alguma forma dizer a que distância as coisas estavam—onde um olho comum veria apenas uma sombra bidimensional do mundo real. Apenas os possuidores dessa habilidade podem mirar com precisão as lendárias armas de distância que matam a distâncias muito além de uma espada, ou usar todo o potencial das conchas de máquinas ultrarrápidas chamadas "carros".

"Visão binocular" seria um termo muito leve para essa habilidade. Só apreciaremos quando tiver um nome devidamente impressionante, como Olhos Místicos da Percepção de Profundidade.

Então, aqui está uma lista de alguns dos meus poderes mágicos favoritos:

- **Telepatia Vibratória.** Ao transmitir vibrações invisíveis pelo próprio ar, dois usuários dessa habilidade podem compartilhar pensamentos. Como resultado, os Telepatas Vibratórios podem formar laços emocionais muito mais profundos do que os possíveis para outros primatas.
- **Traçado Psicométrico.** Ao traçar pequenas linhas finas em uma superfície, o Traçador Psicométrico pode deixar impressões de emoções, história, conhecimento e até a estrutura de outros feitiços. Isso é um nível mais alto do que a Telepatia Vibratória, pois um Traçador Psicométrico pode compartilhar os pensamentos de Traçadores mortos há muito tempo que viveram milhares de anos antes. Ao ler um Traçado e inscrever outro simultaneamente, os Traçadores podem duplicar Traçados; e esses Traçados replicados podem até conter o padrão detalhado de outros feitiços e magias. Assim, os Traçadores empunham um poder quase inimaginável como magos; mas Traçadores podem se meter em problemas ao tentar usar Traçados complicados que eles mesmos não poderiam ter Traçado.
- **Cinética Multidimensional.** Com atos simples, quase automáticos de vontade, os Cinéticos podem fazer com que forças extraordinariamente complexas fluam por meio de pequenos tentáculos e em qualquer objeto físico no alcance do toque—não apenas empurrões, mas combinações de empurrões em muitos pontos que podem efetivamente aplicar torques e torções. A habilidade Cinética é muito mais sutil do que parece à primeira vista: eles a usam não apenas para empunhar objetos existentes com precisão marcial, mas também para aplicar forças que esculpem objetos em formas mais adequadas para empunhadura Cinética. Eles até criam ferramentas que estendem o poder de sua Cinética e lhes permitem esculpir ferramentas cada vez mais finas e complicadas, um ciclo de feedback positivo tão impressionante quanto parece.
- **O Olho.** O usuário dessa habilidade pode perceber torções infinitesimais viajando na Força que liga

a matéria—pequenas vibrações, semelhantes ao poder vital do Sol que incide sobre as folhas, mas muito mais sutis. Um portador do Olho pode sentir objetos muito além do alcance do toque usando as pequenas perturbações que eles fazem na Força. Montanhas a muitos dias de viagem podem ser conhecidas como se estivessem ao alcance do braço. Segundo os portadores do Olho, quando a noite cai e a luz solar falha, eles podem sentir enormes fogos de fusão queimando a distâncias inimagináveis—embora ninguém mais tenha como verificar isso. Diz-se que a posse de um único Olho torna o portador equivalente à realeza.

E finalmente,

- **O Poder Supremo.** O usuário dessa habilidade contém um eco menor e imperfeito de todo o universo, permitindo-lhes buscar caminhos através da probabilidade para qualquer futuro desejado. Se isso parece uma habilidade ridiculamente poderosa, você está certo—a balanceamento de jogo vai por água abaixo com essa. Extremamente raro entre as formas de vida, é o *sekai no ougi* ou "técnica oculta do mundo".

Nada pode se opor ao Poder Supremo exceto o Poder Supremo. Qualquer Poder menos que supremo será simplesmente "compreendido" pelo Supremo e interrompido de alguma forma inconcebível, ou mesmo absorvido na própria base de poder dos Supremos. Por esta razão, o Poder Supremo é às vezes chamado de "técnica mestre das técnicas" ou a "carta trunfo que supera todos os outros trunfos". Os Supremos mais poderosos podem estender sua "compreensão" por distâncias galácticas e eras de tempo, e até mesmo perceber as leis bizarras do "mundo oculto sob o mundo".

Os Supremos foram mortos por imensas catástrofes naturais ou por ataques extremamente rápidos que não lhes dão chance de usar seu poder. Mas todas essas vitórias são, em última análise, uma questão de sorte—não confrontam os Supremos em seu próprio nível de manipulação de probabilidades, e se eles sobreviverem, começarão a dobrar o Tempo para evitar futuros ataques.

Contudo, o próprio Poder Supremo também é perigoso, e muitos Supremos foram destruídos por seus próprios poderes—caindo em uma das falhas em seu eco interno imperfeito do mundo.

Desarmado e trancado em uma cela, um Supremo ainda é uma das formas de vida mais perigosas do planeta. Uma espada pode ser quebrada e um membro pode ser cortado, mas o Poder Supremo é "o poder que não pode ser removido sem remover você."

Talvez porque essa conexão seja tão íntima, os Supremos consideram alguém que perde permanentemente seu Poder Supremo - sem esperança de recuperá-lo - como *schiavo*, ou "morto enquanto respira." Os Supremos argumentam que o Poder Supremo é tão importante a ponto de ser uma parte necessária do que torna uma criatura um fim em si mesma, e não um meio. Os Supremos até insistem que qualquer um que não possua o Poder Supremo não pode começar a compreender verdadeiramente o Poder Supremo, e, portanto, não pode entender por que o Poder Supremo é moralmente importante—um argumento suspeitosamente autosserviente.

Os usuários dessa habilidade formam uma aristocracia absoluta e tratam todas as outras formas de vida como seus peões.

207 — A beleza da ciência consolidada

Os fatos não precisam ser inexplicáveis para serem belos; as verdades não se tornam menos dignas de serem aprendidas se alguém já as conhece; as crenças não se tornam menos valiosas se muitos outros as compartilham…

… e se você só se importar com questões científicas controversas, acabará com a cabeça cheia de lixo.

A mídia acha que apenas a vanguarda da ciência vale a pena ser noticiada. Com que frequência você vê manchetes como "Relatividade Geral Ainda Governa Órbitas Planetárias" ou "Teoria do Flogisto Continua Falsa"? Assim, quando algo se torna ciência sólida, já não é mais uma manchete de última hora. A ciência "digna de notícia" é frequentemente baseada nas evidências mais tênues e está errada metade das vezes — se não estivesse nas fronteiras mais extremas da ciência, não seria notícia de última hora.

As controvérsias científicas são problemas tão difíceis que até pessoas que passaram anos dominando o campo ainda podem se enganar. É isso que gera os acalorados debates que atraem toda a atenção da mídia.

Pior ainda, se você não está no campo e não faz parte do jogo, as controvérsias nem são divertidas.

Ah, claro, você pode se divertir escolhendo um lado em uma discussão. Mas você pode conseguir isso em qualquer jogo de futebol. Não é disso que trata a diversão da ciência.

Ao ler um livro didático bem escrito, você obtém: explicações cuidadosamente formuladas para estudantes iniciantes, matemática derivada passo a passo (quando aplicável), muitos experimentos citados como ilustração (quando aplicável), problemas de teste para você exibir seu novo domínio e uma garantia razoavelmente boa de que o que você está aprendendo é realmente verdadeiro.

Ao ler comunicados de imprensa, você geralmente obtém: explicações falsas que não transmitem nada além da ilusão de compreensão de um resultado que o autor do comunicado não entendeu e que provavelmente tem uma chance maior que 50% de falhar na replicação.

A ciência moderna é construída sobre descobertas, construídas sobre descobertas, construídas sobre descobertas, e assim por diante, até pessoas como Arquimedes, que descobriram fatos como por que os barcos flutuam, que podem fazer sentido mesmo se você não souber sobre outras descobertas. Um bom lugar para começar a percorrer esse caminho é pelo início.

Não fique envergonhado de ler livros didáticos de ciência elementar. Se você quer fingir ser sofisticado, vá encontrar uma peça de teatro para zombar. Se você só quer se divertir, lembre-se que a simplicidade está no cerne da beleza científica.

E pensar que você pode pular direto para a fronteira, quando você não aprendeu a ciência estabelecida, é como…

… como tentar escalar apenas a metade superior do Monte Everest (a única parte que lhe interessa) ficando na base da montanha, dobrando os joelhos e saltando com muita força (para que você possa passar pelas partes chatas).

Agora, não estou dizendo que você nunca deve prestar atenção às controvérsias científicas. Se 40% dos oncologistas acham que meias brancas causam câncer, e os outros 60% discordam violentamente, este é um fato importante de se saber.

Só não vá pensando que a ciência precisa ser controversa para ser interessante.

Ou, aliás, que a ciência precisa ser recente para ser interessante. Uma dieta constante de notícias científicas é ruim para você: você é o que você come, e se você consome apenas reportagens científicas sobre situações fluidas, sem um livro didático sólido ocasionalmente, seu cérebro se transformará em líquido.

208 — Dia da Descoberta Incrível: 1º de Abril

Você deve estar pensando: “1º de abril... não é o Dia da Mentira?”

Sim - e isso proporcionará a cobertura ideal para celebrarmos o Dia da Descoberta Incrível.

Como argumentei em “A Beleza da Ciência Estabelecida”, é um grande problema que a cobertura da mídia sobre ciência se concentre apenas em notícias de última hora. Notícias de última hora, na ciência, ocorrem nos limites mais distantes da fronteira científica, o que significa que a nova descoberta é frequentemente:

- Controversa;
- Apoiada por apenas um experimento;
- Extremamente mais complicada do que um mortal comum pode lidar, e requerendo muito conhecimento científico prévio para entender, razão pela qual não foi resolvida há três séculos;
- Posteriormente, foi demonstrada estar errada.

As pessoas nunca chegam a ver o material sólido, muito menos o material compreensível, porque não é notícia de última hora.

No Dia da Descoberta Incrível, proponho que jornalistas que realmente se preocupam com a ciência possam relatar - sob a cobertura protetora do dia 1º de abril - histórias científicas importantes, mas negligenciadas, como:

- Barcos Explicados: Problema Centenário Resolvido por Nudista em Banheira
- Você Não Deve Cruzar! Esperanças de Turistas em Königsberg Destruídas
- Seus Pulmões Estão em Chamas? Ligação Entre Respiração e Combustão Ganha Aceitação Entre Cientistas

Observe que cada uma dessas manchetes é verdadeira - elas descrevem eventos que, de fato, aconteceram. Eles simplesmente não aconteceram ontem.

Houveram muitas descobertas incríveis e humanamente compreensíveis na história da ciência, que podem ser entendidas sem um doutorado ou mesmo um diploma de bacharel. A palavra operativa aqui é história. Pense no “Eureka!” de Arquimedes quando ele compreendeu a relação entre a água que um navio desloca e a razão pela qual o navio flutua. Isso está longe o suficiente na história científica que você não precisa saber cinquenta outras descobertas para entender a teoria; ela pode ser explicada em alguns gráficos; qualquer um pode ver como é útil; e os experimentos de confirmação podem ser duplicados em sua própria banheira.

A ciência moderna é construída sobre descobertas construídas sobre descobertas construídas sobre descobertas e assim por diante até Arquimedes. Relatar a ciência apenas como notícia de última hora é como entrar em um filme três quartos do caminho, escrever uma história sobre “Homem com as mãos ensanguentadas beija garota segurando arma!” e sair andando novamente.

E se seu editor disser: “Ah, mas nossos leitores não estarão interessados nisso...”

Então, aponte que o Reddit e o Digg não se conectam apenas a notícias de última hora. Eles também

se conectam a páginas da web curtas que fornecem boas explicações da ciência antiga. Os leitores votam positivamente, e isso deve dizer algo. Explique que, se seu jornal não mudar para se parecer mais com o Reddit, você terá que começar a vender drogas para pagar a folha de pagamento. Os editores adoram ouvir esse tipo de coisa, certo?

Na Internet, uma boa nova explicação da ciência antiga é notícia e se espalha como notícia. Por que as seções de ciência dos jornais não poderiam funcionar da mesma maneira? Por que uma nova explicação não vale a pena ser relatada?

Mas tudo isso é muito visionário para um primeiro passo. Por enquanto, vamos apenas ver se algum jornalista por aí pega o Dia da Descoberta Incrível, onde você relata alguma descoberta científica compreensível como se tivesse acabado de ocorrer.

1º de abril. Coloque no seu calendário.

209 — O humanismo é um substituto para a religião?

Por muitos anos antes dos irmãos Wright, as pessoas sonhavam em voar com poções mágicas. Não havia nada de irracional no desejo bruto de voar. Não havia nada de contaminado no desejo de olhar para uma nuvem de cima. Apenas a parte das "poções mágicas" era irracional.

Suponha que você me colocasse em um scanner de ressonância magnética funcional (fMRI) e fizesse um filme da atividade do meu cérebro enquanto eu assistia ao lançamento de um ônibus espacial. (Querer visitar o espaço não é "realista", mas é um sonho essencialmente legítimo — um que pode ser realizado em um universo regido por leis.) O fMRI poderia — talvez sim, talvez não — se assemelhar ao fMRI de um cristão devoto assistindo a uma cena da natividade.

Se um pesquisador obtivesse esse resultado, há muitas pessoas por aí, tanto cristãos quanto alguns ateus, que se regozijariam: "Ha, ha, a viagem espacial é a sua religião!"

Mas isso é traçar a fronteira da categoria errada. É como dizer que, porque algumas pessoas uma vez tentaram voar por meios irracionais, ninguém deveria jamais desfrutar de olhar pela janela de um avião para as nuvens abaixo.

Se um lançamento de foguete é o que é preciso para me dar uma sensação de transcendência estética, não vejo isso como um substituto para a religião. Isso é teomorfismo — o ponto de vista dos religiosos que se regozijam ao assumir que todos que não são religiosos têm um buraco em sua mente que precisa ser preenchido.

Agora, para ser justo com os religiosos, isso não é apenas uma suposição presunçosa. Existem ateus que têm buracos em forma de religião em suas mentes. Eu já vi tentativas de substituir o ateísmo ou até mesmo o transumanismo pela religião. E o resultado é invariavelmente terrível. Completamente terrível. Absoluta e abjetamente terrível.

Chamo esses esforços de "hinos à inexistência de Deus".

Quando alguém se propõe a escrever um hino ateísta — "Salve, oh universo não inteligente", blá, blá, blá — o resultado será, sem exceção, uma porcaria.

Por quê? Porque eles estão sendo imitativos. Porque não têm motivação para escrever o hino, exceto um vago sentimento de que, já que as igrejas têm hinos, eles também deveriam ter um. E, em um nível puramente artístico, isso os coloca muito abaixo da arte religiosa genuína que não é uma imitação de nada, mas uma expressão original de emoção.

Os hinos religiosos foram (frequentemente) escritos por pessoas que sentiam fortemente e escreviam honestamente e se esforçavam seriamente na prosódia e nas imagens de seu trabalho — é isso que dá ao seu trabalho a graça que possui, de integridade artística.

Então, os ateus estão condenados à ausência de hinos?

Existe um teste decisivo para tentativas de pós-teísmo. O teste decisivo é: "Se a religião nunca tivesse existido entre a espécie humana — se nunca tivéssemos cometido o erro original — esta música, esta arte, este ritual, esta forma de pensar ainda faria sentido?"

Se a humanidade nunca tivesse cometido o erro original, não haveria hinos à inexistência de Deus. Mas ainda haveria casamentos, então a noção de uma cerimônia de casamento ateísta faz perfeito sentido — desde que você não comece de repente uma palestra sobre como Deus não existe. Porque, em um mundo onde a religião nunca tivesse existido, ninguém interromperia um casamento para falar sobre a implausibilidade de um conceito hipotético distante. Eles falariam sobre amor, filhos, compromisso, honestidade, devoção, mas quem diabos mencionaria Deus?

E, em um mundo humano onde a religião nunca tivesse existido, ainda haveria pessoas que ficassem com lágrimas nos olhos assistindo ao lançamento de um ônibus espacial.

Por isso, mesmo que um experimento mostre que assistir a um lançamento de ônibus espacial faz as áreas do meu cérebro associadas à "religião" se iluminarem, associadas a sentimentos de transcendência, não vejo isso como um substituto para a religião; espero que as mesmas áreas do cérebro se iluminariam, pelo mesmo motivo, se eu vivesse em um mundo onde a religião nunca tivesse sido inventada.

Um bom "hino ateísta" é simplesmente uma música sobre qualquer coisa que valha a pena cantar e que não seja religiosa.

Além disso, estupidez reversa não é inteligência. O maior idiota do mundo pode dizer que o Sol está brilhando, mas isso não faz com que esteja escuro lá fora. O ponto não é criar uma vida que se assemelhe o menos possível à religião em todos os aspectos superficiais — este é o mesmo tipo de pensamento que inspira hinos à inexistência de Deus. Se a humanidade nunca tivesse cometido o erro original, ninguém estaria tentando evitar coisas que vagamente se assemelhassem à religião. Acredite com precisão e então sinta de acordo: Se os lançamentos espaciais realmente existem, e assistir a um foguete subir faz você querer cantar, então escreva a música, poxa.

Se eu fico com lágrimas nos olhos em um lançamento de ônibus espacial, isso não significa que estou tentando preencher um buraco deixado pela religião — significa que minhas energias emocionais, meu cuidado, estão ligados ao mundo real.

Se Deus falasse claramente e respondesse às orações de forma confiável, Deus se tornaria apenas mais uma coisa entediantemente real, não mais digna de crença do que o carteiro. Se Deus fosse real, isso destruiria a incerteza interna que traz à tona o fervor externo em compensação. E se todos os outros acreditassem que Deus fosse real, isso destruiria a especialidade de ser um dos eleitos.

Se você investir sua energia emocional em viagens espaciais, você não tem essas vulnerabilidades. Posso ver o Ônibus Espacial subir sem perder o assombro. Todos os outros podem acreditar que os Ônibus Espaciais são reais, e isso não os torna menos especiais. Eu não me encurralei em um canto.

A escolha entre Deus e a humanidade não é apenas uma escolha entre duas drogas. Acima de tudo, a humanidade realmente existe.

210 — Escassez

O que se segue é tirado principalmente de “Influência: A Psicologia da Persuasão”, de Robert Cialdini. [1] Tenho três cópias deste livro: uma para mim e duas para emprestar a amigos.

Escassez, como o termo é usado na psicologia social, é quando as coisas se tornam mais desejáveis à medida que parecem menos acessíveis.

- Se você colocar um menino de dois anos em uma sala com dois brinquedos, um brinquedo à vista e o outro atrás de uma parede de Plexiglas, o menino de dois anos ignorará o brinquedo facilmente acessível e irá atrás do aparentemente proibido. Se a parede for baixa o suficiente para ser facilmente escalável, o garoto não será mais propenso a ir atrás de um brinquedo do que do outro. [2]
- Quando o Condado de Dade proibiu o uso ou posse de detergentes com fosfato, muitos residentes de Dade dirigiram até condados vizinhos e compraram grandes quantidades de detergentes para roupas com fosfato. Em comparação com os residentes de Tampa, não afetados pela regulamentação, os residentes de Dade classificaram os detergentes com fosfato como mais suaves, mais eficazes, mais poderosos contra manchas, e até acreditaram que os detergentes com fosfato despejavam mais facilmente. [3]

Da mesma forma, informações que parecem proibidas ou secretas parecem mais importantes e confiáveis:

- Quando os estudantes da Universidade da Carolina do Norte souberam que um discurso contra dormitórios mistos havia sido proibido, eles se tornaram mais contrários aos dormitórios mistos (sem sequer ouvir o discurso). [4]
- Quando um motorista disse que tinha seguro de responsabilidade civil, os jurados experimentais concederam à sua vítima uma média de quatro mil dólares a mais do que se o motorista dissesse que não tinha seguro. Se o juiz informasse posteriormente aos jurados que a informação sobre o seguro era inadmissível e deveria ser ignorada, os jurados concederiam uma média de treze mil dólares a mais do que se o motorista não tivesse seguro. [5]
- Compradores de supermercados, informados por um fornecedor de que a carne bovina estava em oferta limitada, fizeram pedidos do dobro de carne do que compradores informados de que estava prontamente disponível. Compradores informados de que a carne bovina estava em oferta limitada, e, além disso, que a informação sobre a escassez era ela própria escassa—que a escassez não era de conhecimento geral—fizeram pedidos de seis vezes mais carne. (Como o estudo foi realizado em um contexto real, a informação fornecida era de fato correta.) [6]

A teoria convencional para explicar isso é a “reatância psicológica”, termo da psicologia social para “Quando você diz às pessoas que elas não podem fazer algo, elas tentarão ainda mais”. Os instintos fundamentais envolvidos parecem ser a preservação do status e a preservação das opções. Resistimos à dominação, quando qualquer agência humana tenta restringir nossa liberdade. E quando as opções parecem estar em perigo de desaparecer, mesmo por causas naturais, tentamos agarrar a opção antes que ela se vá.

Agarrar-se a opções que estão desaparecendo pode ser uma boa adaptação em uma sociedade de

caçadores-coletores—colher os frutos enquanto continuam maduros—mas em uma sociedade baseada em dinheiro pode ser bastante caro. Cialdini relata que em uma loja de eletrodomésticos que ele observou, um vendedor que via que um cliente estava mostrando sinais de interesse em um eletrodoméstico se aproximava e informava tristemente ao cliente que o item estava fora de estoque, o último havia sido vendido apenas vinte minutos atrás. A escassez criando um salto repentino na desejabilidade, o cliente frequentemente perguntava se havia alguma chance de o vendedor localizar um item não vendido no depósito, armazém ou em qualquer lugar. “Bem,” dizia o vendedor, “isso é possível, e estou disposto a verificar; mas entendo que este é o modelo que você quer, e se eu encontrar a este preço, você o levará?”

Como Cialdini destaca, um sinal principal desse mau funcionamento é que você sonha em possuir algo, em vez de usá-lo. (Timothy Ferriss oferece um conselho semelhante sobre planejar sua vida: pergunte quais experiências contínuas o fariam feliz, em vez de quais posses ou mudanças de status.)

Mas o problema realmente fundamental de desejar o inalcançável é que, assim que você realmente o consegue, ele deixa de ser inalcançável. Se não podemos nos alegrar com o meramente disponível, nossas vidas estarão sempre frustradas...

Referências

[1] Robert B. Cialdini, Influence: The Psychology of Persuasion: Revised Edition (New York: Quill, 1993).

[2] Sharon S. Brehm and Marsha Weintraub, “Physical Barriers and Psychological Reactance: Two- year--olds’ Responses to Threats to Freedom,” Journal of Personality and Social Psychology 35 (1977): 830–836.

[3] Michael B. Mazis, Robert B. Settle, and Dennis C. Leslie, “Elimination of Phosphate Detergents and Psychological Reactance,” Journal of Marketing Research 10 (1973): 2; Michael B. Mazis, “Antipollution Measures and Psychological Reactance Theory: A Field Experiment,” Journal of Personality and Social Psychology 31 (1975): 654–666.

[4] Richard D. Ashmore, Vasantha Ramchandra, and Russell A. Jones, “Censorship as an Attitude Change Induction,” Paper presented at Eastern Psychological Association meeting (1971).

[5] Dale Broeder, “The University of Chicago Jury Project,” Nebraska Law Review 38 (1959): 760–774.

[6] A. Knishinsky, “The Effects of Scarcity of Material and Exclusivity of Information on Industrial Buyer Perceived Risk in Provoking a Purchase Decision” (Doctoral dissertation, Arizona State University, 1982).

211 — O Sagrado Mundano

Estava lendo (a primeira metade de) *The Constant Fire* (O fogo constante), de Adam Frank [1], em preparação para meu diálogo com ele no Bloggingheads. O livro de Adam Frank é sobre a experiência do sagrado. Eu normalmente não a chamaria assim, mas, claro, conheço a experiência da qual Frank está falando. É o que sinto quando assisto a um vídeo do lançamento de um ônibus espacial; ou o que sinto - em menor grau, porque neste mundo é muito comum - quando olho para as estrelas à noite e penso sobre o que elas significam. Ou o nascimento de uma criança, digamos. Aquilo que é significativo na História em Desdobramento.

Adam Frank sustenta que essa experiência é algo que a ciência tem profundamente em comum com a religião. Em oposição a, por exemplo, ser uma qualidade humana básica que a religião corrompe.

The Constant Fire cita "As Variedades da Experiência Religiosa" de William James, dizendo:

> Religião... deve significar para nós os sentimentos, atos e experiências de homens individuais em sua solidão; na medida em que eles se apreendem em relação ao que quer que considerem divino[26].

E este tema é desenvolvido ainda mais: a sacralidade é algo intensamente privado e individual.

O que me deixou completamente perplexo. Devo não ter nenhum sentimento de sacralidade se sou uma das muitas pessoas assistindo ao vídeo da SpaceShipOne ganhando o X-Prize? Por que não? Devo pensar que minha experiência de sacralidade tem que ser de alguma forma diferente da de todas as outras pessoas assistindo? Por que, quando todos temos o mesmo design cerebral? De fato, por que eu precisaria acreditar que sou único? (Mas "único" é outra palavra que Adam Frank usa; a "experiência única do sagrado" de fulano de tal.) O sentimento é privado no mesmo sentido em que temos dificuldade em comunicar qualquer experiência? Então por que enfatizar isso em relação à sacralidade, em vez de espirrar?

A luz se acendeu quando percebi estar olhando para um truque da Epistemologia do Lado Negro - se você torna algo privado, isso o protege de críticas. Você pode dizer: "Você não pode me criticar, porque esta é minha experiência interior privada, à qual você nunca pode ter acesso para questioná-la."

Mas o preço de se proteger de críticas é que você é lançado na solidão - a solidão que William James admirava como o núcleo da experiência religiosa, como se a solidão fosse uma coisa boa.

Tais relíquias da Epistemologia do Lado Negro são a chave para entender as muitas maneiras pelas quais a religião distorce a experiência do sagrado:

Misteriosidade - por que o sagrado tem que ser misterioso? Um lançamento de ônibus espacial se sai muito bem sem ser misterioso. Quanto menos eu apreciaria as estrelas se não soubesse o que elas são, se fossem apenas pequenos pontos no céu noturno? Mas se suas crenças religiosas são questionadas - se alguém perguntar: "Por que Deus não cura amputados?" - Então você se refugia e diz, em um tom de profunda profundidade: "É um mistério sagrado!" Há perguntas que não devem ser feitas e respostas que não devem ser reconhecidas, para defender a mentira. Assim, a impossibilidade de responder passa a ser associada à sacralidade. E o preço de se proteger de críticas é desistir da verdadeira curiosidade que realmente deseja

26 NT. Texto original em inglês. *Religion . . . shall mean for us the feelings, acts, and experiences of individual men in their solitude; so far as they apprehend themselves to stand in relation to whatever they may consider the divine.*

encontrar respostas. Você adorará sua própria ignorância das perguntas temporariamente não respondidas de sua própria geração — provavelmente incluindo aquelas que já foram respondidas.

Fé — nos primórdios da religião, quando as pessoas eram mais ingênuas, quando até mesmo pessoas inteligentes realmente acreditavam nessas coisas, as religiões apostavam sua reputação no testemunho de milagres em suas escrituras. E os arqueólogos cristãos partiram realmente esperando encontrar as ruínas da Arca de Noé. Mas quando nenhuma evidência apareceu, então a religião executou o que William Hartley chamou de recuo para o compromisso: "Eu acredito porque acredito!" Assim, a crença sem boas evidências passou a ser associada à experiência do sagrado. E o preço de se proteger de críticas é que você sacrifica sua capacidade de pensar claramente sobre aquilo que é sagrado, de progredir em sua compreensão do sagrado e de abandonar os erros.

Experimentalismo — se antes você pensava que o arco-íris era um contrato sagrado de Deus com a humanidade, e então você começa a perceber que Deus não existe, então você pode fazer um recuo em direção da experiência pura - para se elogiar apenas por sentir sensações tão maravilhosas quando você pensa em Deus, quer Deus realmente exista ou não. E o preço de se proteger de críticas é o solipsismo: sua experiência é despojada de seus referentes. Que terrível sentimento de vazio seria assistir a um ônibus espacial subindo em um pilar de chamas e dizer a si mesmo: "Mas não importa se o ônibus espacial realmente existe, contanto que eu sinta."

Separação — se o reino sagrado não está sujeito às regras ordinárias de evidência ou investigável por meios ordinários, então ele deve ser diferente em espécie do mundo da matéria mundana: e assim somos menos propensos a pensar em um ônibus espacial como um candidato à sacralidade, porque é um trabalho de meras mãos humanas. Keats perdeu sua admiração pelo arco-íris e o rebaixou ao "catálogo enfadonho de coisas mundanas" pelo crime de sua trama e textura serem conhecidas. E o preço de se proteger de todas as críticas ordinárias é que você perde a sacralidade de todas as coisas meramente reais.

Privacidade - sobre isso já falei.

Tais distorções são a razão pela qual é melhor não tentarmos salvar a religião. Não, nem mesmo na forma de "espiritualidade". Tire as instituições e os erros factuais, subtraia as igrejas e as escrituras, e você fica com… todo esse absurdo sobre misteriosidade, fé, experiência solipsista, solidão privada e descontinuidade.

A mentira original é apenas o começo do problema. Então você tem todos os maus hábitos de pensamento que evoluíram para defendê-la. A religião é um cálice envenenado, do qual é melhor nem mesmo bebermos. A espiritualidade é o mesmo cálice depois que a pelota original de veneno foi retirado, e apenas a porção dissolvida permanece - um pouco menos letal, mas ainda assim não é bom para você.

Quando uma mentira foi defendida por eras e eras, a verdadeira origem dos hábitos herdados se perde nas brumas, com camada após camada de doença não documentada; então os sábios, penso eu, começarão do zero, em vez de tentar descartar seletivamente a mentira original enquanto mantêm os hábitos de pensamento que a protegeram. Apenas admita que você estava errado, desista inteiramente do erro, pare de defendê-lo, pare de tentar dizer que você estava até um pouco certo, pare de tentar salvar as aparências, apenas diga "Opa!" e jogue fora a coisa toda e comece de novo.

Essa capacidade - de realmente, realmente, sem defesa, admitir que você estava totalmente errado - é a razão pela qual a experiência religiosa nunca será como a experiência científica. Nenhuma religião pode absorver essa capacidade sem se perder inteiramente e se tornar simples humanidade...

… Para apenas olhar para as estrelas distantes. Acreditável sem esforço, sem uma luta constante e perturbadora para afastar sua consciência da contra-evidência. Verdadeiramente lá no mundo, a experiência unida ao referente, uma parte sólida daquela história em desenvolvimento. Conhecível sem ameaça, oferecendo verdadeiro alimento para a curiosidade. Compartilhada em união com os muitos outros espectadores, sem necessidade de recuar para a privacidade. Feita do mesmo tecido que você e todas as outras coisas. O mais sagrado e belo, o sagrado mundano.

Referências

[1] Adam Frank, The Constant Fire: Beyond the Science vs. Religion Debate (University of California Press, 2009).

212 — Para divulgar a ciência, mantenha-a em segredo

Às vezes, me pergunto se os pitagóricos não tinham razão.

Sim, já escrevi sobre como a "ciência" é inerentemente pública. Argumentei que a "ciência" se distingue do conhecimento meramente racional pela capacidade, em princípio, de reproduzir experimentos científicos por conta própria, de saber sem depender de autoridades. Afirmei que a "ciência" deveria ser definida como o conhecimento publicamente acessível da humanidade. Até sugeri que as gerações futuras considerarão todos os artigos não publicados em periódicos de acesso aberto como não-científicos, isto é, não pode fazer parte do conhecimento público da humanidade se você cobra para que as pessoas o leiam.

Mas essa é apenas uma visão do futuro. Em outra visão, o conhecimento que hoje chamamos de "ciência" é retirado do domínio público — os livros e periódicos são escondidos, guardados por cultos místicos de gurus vestindo túnicas, exigindo rituais de iniciação temíveis para ter acesso — de modo que mais pessoas realmente o estudem.

Quero dizer, agora mesmo, as pessoas podem estudar ciência, mas não o fazem.

"Escassez" é como isso é chamado na psicologia social. O que parece estar em oferta limitada é mais valorizado. E esse efeito é especialmente forte com informações — somos muito mais propensos a tentar obter informações que acreditamos serem secretas e a valorizá-las mais quando as obtemos.

Com a ciência, eu acredito, as pessoas presumem que se a informação está livremente disponível, ela não deve ser importante. Então, em vez disso, as pessoas se juntam a cultos que têm o bom senso de manter suas Grandes Verdades em segredo. A Grande Verdade pode ser, na realidade, um monte de besteira, mas é mais satisfatória do que a ciência coerente, porque é secreta.

A ciência é a grande Carta Roubada dos nossos tempos, deixada à vista de todos e ignorada.

Claro, a abertura científica ajuda a elite científica. Eles já passaram pelos rituais de iniciação. Mas para o resto do planeta, a ciência é mantida em segredo cem vezes mais efetivamente ao torná-la livremente disponível do que se seus livros fossem guardados em cofres e você tivesse que caminhar sobre carvão em brasa para ter acesso. (Sendo este um teste realmente temível, já que os grandes segredos do isolamento térmico só estão disponíveis para os Físicos-Iniciados do Terceiro Nível.)

Se o conhecimento científico estivesse escondido em antigos cofres (em vez de escondido em inconvenientes periódicos pagos), pelo menos então as pessoas tentariam entrar nos cofres. Elas estariam desesperadas para aprender ciência. Especialmente quando vissem o poder que os Físicos do Oitavo Nível poderiam exercer e fossem informadas de que não tinham permissão para saber a explicação.

E se você tentasse iniciar um culto em torno de, digamos, cientologia, você obteria algum grau de interesse público, no começo. Mas as pessoas muito rapidamente começariam a fazer perguntas incômodas como "Por que você não deu uma demonstração pública dos seus poderes de Oitavo Nível, como os Físicos?" e "Como nenhum dos Mestres Matemáticos parece querer se juntar à sua seita?" e "Por que eu deveria seguir seu Fundador quando ele não é um Oitavo Nível de nada fora do próprio culto?" e "Por que eu deveria estudar seu culto primeiro, quando os Dentistas do Destino podem fazer coisas que são muito mais impressionantes?"

Quando você olha por essa perspectiva, o escape da matemática do culto pitagórico começa a parecer um grande erro estratégico para a humanidade.

Agora, eu sei o que você vai dizer: "Mas a ciência é cercada por rituais de iniciação temíveis! Além disso, é inerentemente difícil de aprender! Por que isso não conta?" Porque o público pensa que a ciência está livremente disponível, é por isso. Se você tem permissão para aprender, não deve ser importante o suficiente para ser aprendido.

É um problema de imagem, as pessoas tomando suas dicas a partir das atitudes dos outros. Qualquer um pode entrar no supermercado e comprar uma lâmpada, e ninguém olha para ela com admiração e reverência. A física supostamente não é secreta (mesmo que você não saiba), e há uma explicação de um parágrafo no jornal que soa vagamente autoritária e convincente — essencialmente, ninguém trata a lâmpada como um mistério sagrado, então você também não.

Até as coisas mais simples, objetos completamente inertes como crucifixos, podem se tornar mágicos se todos olharem para eles como se fossem mágicos. Mas como você teoricamente tem permissão para saber por que a lâmpada funciona sem escalar a montanha para encontrar o remoto Monastério dos Eletricistas, não há necessidade de realmente se incomodar em aprender.

Agora, como a ciência de fato tem rituais de iniciação, tanto sociais quanto cognitivos, os cientistas não estão totalmente insatisfeitos com sua ciência. O problema é que, no mundo atual, muito poucas pessoas se incomodam em estudar ciência em primeiro lugar. A ciência não pode ser o verdadeiro Conhecimento Secreto, porque qualquer um tem permissão para conhecê-la — mesmo que, na verdade, eles não o façam.

Se o Grande Segredo da Seleção Natural, transmitido por Darwin Que Não É Esquecido, só fosse revelado a você depois que você pagasse US$ 2.000 e passasse por uma cerimônia envolvendo tochas, túnicas, máscaras e o sacrifício de um boi, então quando lhe mostrassem os fósseis, e mostrassem o cabo óptico passando pela retina sob um microscópio, e finalmente lhe contassem a Verdade, você diria "Essa é a coisa mais brilhante de todos os tempos!" e ficaria satisfeito. Depois disso, se algum outro culto tentasse lhe dizer que, na verdade, foi um homem barbudo no céu há 6.000 anos, você riria muito.

E sabe de uma coisa, poderia até ser mais divertido fazer as coisas dessa maneira. Especialmente se a iniciação exigisse que você juntasse algumas das evidências por conta própria — sozinho ou com colegas de classe — antes que você pudesse dizer ao seu Sensei de Ciência que estava pronto para avançar para o próximo nível. Não seria eficiente, claro, mas seria divertido.

Se a humanidade nunca tivesse cometido o erro — nunca tivesse seguido o caminho religioso e nunca tivesse aprendido a temer qualquer coisa que lembre religião — então talvez a cerimônia de concessão do título de doutor envolvesse litanias e cânticos, porque, ei, é disso que as pessoas gostam. Por que tirar a diversão de tudo?

Talvez estejamos apenas fazendo isso errado.

E não, não estou propondo seriamente que tentemos reverter os últimos quinhentos anos de abertura e classificar toda a ciência como secreta. Pelo menos, não no momento. A eficiência é importante por enquanto, especialmente em coisas como pesquisa médica. Estou apenas explicando por que é que não contarei a ninguém o Segredo de como a inefável diferença entre o azul e o vermelho surge de meros átomos por menos de US$ 100.000—

Ahem! Eu quis dizer, estou lhe contando sobre essa visão de uma Terra alternativa, para que você dê à ciência o mesmo tratamento que dá aos cultos. Para que você não desvalorize a verdade científica quando a aprender, só porque ela não parece estar protegida apropriadamente ao seu valor. Imagine as túnicas e máscaras. Visualize-se rastejando para dentro dos cofres e roubando o Conhecimento Perdido de Newton. E não se deixe enganar por nenhuma organização que realmente use túnicas e máscaras, a menos que eles também lhe mostrem os dados.

As pessoas parecem ter buracos em suas mentes para o Conhecimento Esotérico, os Segredos Profundos, a Verdade Oculta. E eu nem estou criticando essa psicologia! Existem verdades ocultas esotéricas

profundas e secretas, como a mecânica quântica ou a estrutura bayesiana. Nós apenas nos acostumamos a apresentar a Verdade Oculta de uma maneira muito insatisfatória, envolta em falsa mundanidade.

Mas se os buracos para o conhecimento secreto não forem preenchidos por crenças verdadeiras, serão preenchidos por crenças falsas. Não há nada além da ciência para aprender — a energia emocional deve ser investida na realidade, ou desperdiçada em total absurdo, ou destruída. Para mim, acho que é melhor investir a energia emocional; a diversão não deve ser desnecessariamente descartada.

Agora mesmo, temos o pior dos dois mundos. A ciência não é realmente gratuita, porque os cursos são caros e os livros didáticos são caros. Mas o público pensa que qualquer um tem permissão para saber, então não deve ser importante.

Idealmente, você gostaria de arranjar as coisas do outro jeito.

213 — Cerimônia de iniciação

As tochas que iluminavam a escada estreita queimavam intensamente e na cor errada, chamas como ouro derretido ou sóis fragmentados.

192... 193...

As sandálias de Brennan clicavam suavemente nos degraus de pedra, soando em sequência, como dominós caindo muito lentamente.

227... 228...

Meia volta à sua frente, uma franja de tecido escuro descia sussurrando as escadas, a figura encapuzada permanecendo fora de vista.

239... 240...

Não falta muito mais, previu Brennan para si mesmo, e sua previsão estava correta: dezesseis vezes dezesseis degraus era o número, e eles estavam diante do portal de vidro.

O grande portão curvo havia sido forjado com astúcia, humor e muita atenção aos índices de refração: ele distorcia a luz, dobrava-a, refletia-a e geralmente a abusava, de modo que havia sugestões do que estava do outro lado (fontes de luz mais fortes, paredes escuras), mas nenhuma maneira possível de ver através—exceto, é claro, se você tivesse a chave: a contraporta, espessa por fina e fina por espessa, caso em que as duas se cancelariam.

Da figura encapuzada ao lado de Brennan, duas mãos emergiram, enluvadas em tecido reflexivo para ocultar a cor da pele. Dedos como pequenos espelhos agarraram as alças do portão distorcido—alças que Brennan não havia adivinhado; em toda aquela distorção, as formas só podiam ser antecipadas, não vistas.

“Você quer saber?” sussurrou o guia; um sussurro quase tão alto quanto uma voz comum, mas sem revelar o menor indício de gênero.

Brennan hesitou. A resposta à pergunta parecia suspeita, de fato, extraordinariamente óbvia, mesmo para um ritual.

“Sim,” disse Brennan finalmente.

O guia apenas o observava em silêncio.

“Sim, eu quero saber,” disse Brennan.

“Saber o quê, exatamente?” sussurrou a figura.

A face de Brennan se contorceu em concentração, tentando visualizar o jogo até o fim, e esperando não ter estragado tudo já; até que finalmente ele recorreu ao primeiro e último recurso, a qual é a verdade:

“Não importa,” disse Brennan, “a resposta ainda é sim.”

O portão de vidro se abriu ao meio e deslizou, com apenas o menor som de raspagem, para a pedra circundante.

A sala revelada estava forrada, de parede a parede, com figuras encapuzadas em tecido que absorvia a luz. As paredes retas não eram de pedra negra, mas espelhadas, formando uma grade quadrada de vestes escuras até o infinito em todas as direções; de modo que parecia como se as pessoas de uma cidade muito maior, ou talvez toda a humanidade, observassem em assembleia. Havia um leve calor úmido no ar da sala, o fôlego dos reunidos: um cheiro de multidão.

O guia de Brennan se moveu para o centro da praça, onde queimavam quatro tochas daquela chama amarela implacável. Brennan o seguiu, e quando parou, percebeu com um leve choque que todos os capuzes agora olhavam diretamente para ele. Brennan jamais em sua vida havia sido o foco de uma atenção tão absoluta; era assustador, mas não totalmente desagradável.

"Ele está aqui," disse o guia naquele estranho sussurro alto.

A infinita grade de figuras encapuzadas respondeu em uma só voz: perfeitamente combinada, exatamente sincronizada, de modo que nenhum indivíduo pudesse ser distinguido dos demais, e traído:

"Quem está ausente?"

"Jakob Bernoulli," entoou o guia, e as paredes responderam:

"Está morto, mas não esquecido."

"Abraham de Moivre,"

"Está morto, mas não esquecido."

"Pierre-Simon Laplace,"

"Está morto, mas não esquecido."

"Edwin Thompson Jaynes,"

"Está morto, mas não esquecido."

"Eles morreram," disse o guia, "e estão perdidos para nós; mas ainda temos uns aos outros, e o projeto continua."

No silêncio, o guia se virou para Brennan e estendeu a mão, na qual descansava um pequeno anel de material quase transparente.

Brennan deu um passo à frente para pegar o anel—

Mas a mão se fechou firmemente.

"Se três quartos dos humanos nesta sala são mulheres," disse o guia, "e três quartos das mulheres e metade dos homens pertencem à Heresia da Virtude, e eu sou um Virtuista, qual é a probabilidade de eu ser um homem?"

"Dois-onze avos," disse Brennan confiantemente.

Houve um momento de silêncio absoluto.

Então uma risada chocada.

O sussurro do guia veio novamente, realmente quieto, desta vez, quase inexistente:

"Na verdade, é um sexto."

As bochechas de Brennan estavam tão vermelhas que ele pensou que seu rosto poderia derreter. O instinto era muito forte para sair correndo da sala e subir as escadas e fugir da cidade e mudar seu nome e começar sua vida de novo e acertar desta vez.

"Um erro honesto é pelo menos honesto," disse o guia, mais alto agora, "e podemos conhecer a honestidade por sua renúncia. Se sou um Virtuista, qual é a probabilidade de eu ser um homem?"

"Um—" Brennan começou a dizer.

Então ele parou. Novamente, o silêncio horrível.

"Apenas diga 'um sexto' já," sussurrou a figura, desta vez alto o suficiente para as paredes ouvirem; então houve mais risos, nem todos gentis.

Brennan estava respirando rapidamente e havia suor em sua testa. Se ele estivesse errado sobre isso, ele realmente iria fugir da cidade. "Três quartos mulheres vezes três quartos Virtuistas são nove dezesseis avos de Virtuistas femininas nesta sala. Um quarto homens vezes metade dos Virtuistas são dois dezesseis avos de Virtuistas masculinos. Se eu tiver apenas essa informação e o fato de que você é um Virtuista, eu então estimaria chances de dois para nove, ou uma probabilidade de dois-onze avos, que você é masculino. Embora eu não acredite, de fato, que a informação dada esteja correta. Por um lado, parece muito arrumado. Por outro, há um número ímpar de pessoas nesta sala."

A mão se estendeu novamente e se abriu.

Brennan pegou o anel. Parecia quase invisível, à luz da tocha; não era de vidro, mas de algum material com um índice de refração muito próximo ao do ar. O anel estava quente da mão do guia, e parecia uma coisa viva minúscula enquanto abraçava seu dedo.

O alívio foi tão grande que ele quase não ouviu as figuras encapuzadas aplaudindo.

Do guia encapuzado veio um último sussurro:

"Você agora é um noviço da Conspiração Bayesiana."

Parte R - Fisicalismo 201

214 - Mão vs. Dedos

Voltando ao nosso tópico original: Reducionismo e a Falácia da Projeção Mental.

Pode haver problemas emocionais em aceitar o reducionismo, se você acha que as coisas precisam ser fundamentais para serem divertidas. Mas essa posição nos leva a nunca ter alegria em nada mais complicado do que um quark, e por isso prefiro rejeitá-la.

Para recapitular, a tese reducionista é que usamos modelos de múltiplos níveis por razões computacionais, mas a realidade física tem apenas um único nível.

Gostaria de apresentar o seguinte enigma: quando você pega um copo de água, é a sua mão que o pega?

A maioria das pessoas, claro, escolhe a resposta popular ingênua: "Sim".

Recentemente, no entanto, os cientistas fizeram uma descoberta impressionante: não é a sua mão que segura o copo, são, na verdade, seus dedos, polegar e palma.

Sim, eu sei! Eu também fiquei chocado. Mas parece que depois que os cientistas mediram as forças exercidas no copo por cada um dos seus dedos, seu polegar e sua palma, eles descobriram que não havia nenhuma força restante - então a força exercida pela sua mão deve ser zero.

O tema aqui é que, se você puder ver como (não apenas saber que) um nível superior se reduz a um nível inferior, eles não parecerão coisas separadas dentro do seu mapa; você poderá ver como é bobo pensar que seus dedos poderiam estar em um lugar e sua mão em outro; você poderá ver como é bobo discutir se é a sua mão que pega o copo ou seus dedos.

A palavra operativa é "ver", como em visualização concreta. Imaginar sua mão faz você imaginar os dedos, o polegar e a palma; inversamente, imaginar os dedos, o polegar e a palma faz você identificar uma mão na imagem mental. Assim, o nível alto do seu mapa e o nível baixo do seu mapa estarão fortemente ligados em sua mente.

Na realidade, é claro, os níveis estão ligados ainda mais do que isso - ligados pela ligação mais forte possível: a identidade física. Você pode ver isso: você pode ver que dizer (1) "mão" ou (2) "dedos, polegar e palma" não se refere a coisas diferentes, mas a diferentes pontos de vista.

Mas suponha que você não tenha o conhecimento para vincular tão fortemente os níveis do seu mapa. Por exemplo, você poderia ter um "scanner de mão" que mostrasse uma "mão" como um ponto em um mapa (como um antigo visor de radar) e scanners semelhantes para dedos/polegares/palmas; então você veria um aglomerado de pontos ao redor da mão, mas você conseguiria imaginar o ponto da mão se afastando dos outros. Assim, mesmo que a realidade física da mão (isto é, a coisa à qual o ponto corresponde) fosse idêntica/estritamente composta das realidades físicas dos dedos, polegar e palma, você seria incapaz de ver esse fato; mesmo que alguém lhe dissesse, ou você adivinhasse pela correspondência dos pontos, você apenas saberia o fato da redução, não o veria. Você ainda conseguiria imaginar o ponto da mão se movendo independentemente, mesmo que, se a composição física dos sensores fosse mantida constante, seria fisicamente impossível que isso realmente acontecesse.

Ou, em um nível ainda mais baixo de ligação, as pessoas poderiam simplesmente lhe dizer “Há uma mão ali e alguns dedos ali” - nesse caso, você saberia pouco mais do que uma IA à Moda Antiga representando a situação usando tokens LISP sugestivamente nomeados. Não haveria nada obviamente contraditório em afirmar:

⊢ Dentro(Sala, Mão)

⊢ ¬ Dentro(Sala, Dedos),

porque você não possuiria o conhecimento

⊢ Dentro(x, Mão)) ⇒ Dentro(x, Dedos).

Nada disso diz que uma mão pode realmente separar sua existência de seus dedos e rastejar, como um fantasma, pela sala; apenas diz que uma IA à Moda Antiga com uma representação proposicional pode não saber mais do que isso. O mapa não é o território.

Em particular, você não deve tirar muitas conclusões de como parece conceitualmente possível, na mente de algum pensador específico, separar a mão de seus elementos constituintes de dedos, polegar e palma. A possibilidade conceitual não é o mesmo que a possibilidade lógica ou a possibilidade física.

É conceitualmente possível para você que 235.757 seja primo, porque você não sabe mais do que isso. Mas não é logicamente possível que 235.757 seja primo; se você fosse logicamente onisciente, 235.757 seria obviamente composto (e você saberia os fatores). É por isso que temos a noção de mundos possíveis impossíveis, para podermos colocar distribuições de probabilidade em proposições que podem ou não ser de fato logicamente impossíveis.

E você pode imaginar filósofos que criticam os “eliminativistas dos dedos”, que contradizem os fatos diretos da experiência - podemos sentir nossa mão segurando o copo, afinal - sugerindo que as “mãos” não existem realmente, caso em que, obviamente, o copo cairia. E filósofos que sugerem “leis de ponte apendicular” para explicar como uma configuração particular de dedos evoca uma mão à existência - com a observação, é claro, de que enquanto nosso mundo contém essas leis de ligação apendicular particulares, as leis poderiam ter sido concebivelmente diferentes e, portanto, não são em nenhum sentido fatos necessários, etc.

Todos esses são casos da Falácia da Projeção Mental e do que chamo de “realismo filosófico ingênuo” - a confusão de intuições filosóficas com informações diretas e verdadeiras sobre a realidade. Sua incapacidade de imaginar algo é apenas um fato computacional sobre o que seu cérebro pode ou não imaginar. Outro cérebro pode funcionar de maneira diferente.

215 - Átomos zangados

A física fundamental — quarks e coisas do tipo — está muito distante dos níveis que podemos ver, como mãos e dedos. Na melhor das hipóteses, você pode saber como replicar os experimentos que mostram que sua mão (como tudo o mais) é composta de quarks, e você pode saber como derivar algumas equações para coisas como átomos, nuvens de elétrons e moléculas.

Na pior das hipóteses, a existência de quarks sob sua mão pode ser apenas algo que lhe disseram. Nesse caso, é questionável em que sentido você pode dizer que "sabe" disso, mesmo que repita a mesma palavra "quark" que um físico usaria para transmitir conhecimento a outro físico.

De qualquer forma, você não pode realmente ver a identidade entre os níveis — ninguém tem um cérebro grande o suficiente para visualizar avogadros de quarks e reconhecer um padrão de mão neles.

Mas pelo menos entendemos o que as mãos fazem. Mãos empurram coisas, exercem forças sobre elas. Quando nos falam sobre átomos, visualizamos pequenas bolas de bilhar colidindo umas com as outras. Isso faz parecer óbvio que os "átomos" também podem empurrar as coisas, batendo nelas.

Agora, essa noção de átomos não está totalmente correta. Mas no que diz respeito à imaginação humana, é relativamente fácil imaginar nossa mão sendo composta por uma pequena galáxia de bolas de bilhar rodopiantes, empurrando coisas quando nossos "dedos" as tocam. Demócrito imaginou isso há 2.400 anos, e houve um tempo, aproximadamente entre 1803 e 1922, em que a Ciência pensou que ele estava certo.

Mas e quanto à, digamos, raiva?

Como pequenas bolas de bilhar poderiam ficar zangadas? Pequenas carinhas franzidas nas bolas de bilhar?

Coloque-se no lugar de, digamos, um caçador-coletor — alguém que pode nem mesmo ter uma noção de escrita, muito menos a noção de usar matéria básica para realizar computações — alguém que não tem ideia de que existem coisas como neurônios. Então você pode imaginar o abismo funcional que seus ancestrais podem ter percebido entre bolas de bilhar e "Grrr! Aaarg!"

Esqueça a experiência subjetiva por um momento e considere a pura lacuna comportamental entre a raiva e as bolas de bilhar. A diferença entre o que pequenas bolas de bilhar fazem e o que a raiva faz as pessoas fazerem. A raiva pode fazer as pessoas levantarem os punhos e baterem em alguém — ou dizerem coisas maldosas pelas costas — ou plantar escorpiões em suas tendas à noite. Bolas de bilhar apenas empurram as coisas.

Tente se colocar no lugar do caçador-coletor que nunca teve o "Eureka!" do processamento de informações. Tente evitar o viés retrospectivo sobre coisas como neurônios e computadores. Só então você conseguirá ver o abismo explicativo intransponível:

Como você pode explicar o comportamento zangado em termos de bolas de bilhar?

Bem, a conjectura materialista óbvia é que as pequenas bolas de bilhar empurram seu braço e fazem você bater em alguém, ou empurram sua língua para que insultos saiam.

Mas como as pequenas bolas de bilhar sabem como fazer isso — ou como guiar sua língua e dedos por meio de planos de longo prazo — se elas próprias não estão zangadas?

E, além disso, se você não for seduzido pelo — suspiro! — cientificismo, você pode ver de uma perspectiva de primeira pessoa que essa explicação é obviamente falsa. Os átomos podem empurrar seu braço, mas não podem fazer você querer nada.

Alguém pode apontar que beber vinho pode deixá-lo zangado. Mas quem diz que o vinho é feito exclusivamente de pequenas bolas de bilhar? Talvez o vinho apenas contenha uma potência de raiva.

Claramente, o reducionismo é apenas uma noção falha.

(O noviço se desvia e diz "A arte me falhou"; o mestre se desvia e diz "Eu falhei com minha arte.")

O que é preciso para cruzar esse abismo? Não é apenas a ideia de "neurônios" que "processam informações" — se você disser apenas isso e nada mais, isso apenas insere uma regra mágica e inexplicada de cruzamento de nível em seu modelo, aonde você vai de bolas de bilhar a pensamentos.

Mas um programador de Inteligência Artificial que sabe como criar um programa de xadrez a partir da matéria básica deu um passo genuíno para cruzar o abismo. Se você entender conceitos como consequencialismo, encadeamento reverso, funções de utilidade e árvores de busca, você pode fazer sistemas meramente causais/mecânicos computarem planos.

O truque é mais ou menos assim: para cada possível movimento de xadrez, compute os movimentos que seu oponente poderia fazer, depois suas respostas a esses movimentos, e assim por diante; avalie a posição mais distante que você pode ver usando algum algoritmo local (você pode simplesmente contar o material); então trace de volta usando minimax para encontrar o melhor movimento no tabuleiro atual; então faça esse movimento.

Em termos mais amplos, se você tiver cadeias de causalidade na mente que tenham um tipo de mapeamento - um espelho, um eco - para o que ocorre no ambiente, então você pode executar uma função de utilidade sobre os resultados da imaginação e encontrar uma ação que alcance um objetivo altamente classificado pela função de utilidade e executar essa ação. Não é necessário que as cadeias de causalidade na mente, que são semelhantes ao ambiente, sejam feitas de bolas de bilhar que têm pequenas auras de intencionalidade. Os transistores do Deep Blue não precisam ter pequenas peças de xadrez esculpidas neles para funcionar. Veja também The Simple Truth.

Tudo isso ainda é tremendamente simplificado, mas deve, pelo menos, reduzir o comprimento aparente do abismo. Se você consegue entender tudo isso, pode ver como um planejador construído a partir da matéria básica pode ser influenciado pelo álcool para produzir mais comportamentos raivosos. As bolas de bilhar no álcool empurram as bolas de bilhar que compõem a função de utilidade.

Mas mesmo que você saiba como escrever pequenas IAs, você não pode visualizar a transição de nível entre transistores e xadrez. Há transistores demais e movimentos demais para verificar.

Da mesma forma, mesmo se você soubesse todos os fatos da neurologia, você seria incapaz de visualizar a transição de nível entre neurônios e raiva — muito menos a transição de nível entre átomos e raiva. Não da maneira como você pode visualizar uma mão consistindo em dedos, polegar e palma.

E suponha que um cientista cognitivo lhe diga categoricamente "A raiva é hormônio"? Mesmo que você repita as palavras, isso não significa que você cruzou o abismo. Você pode acreditar que acredita nisso, mas isso não é o mesmo que entender o que pequenas bolas de bilhar têm a ver com querer bater em alguém.

Então você surge com interpretações como, "A raiva é um mero hormônio, é causada por pequenas moléculas, então não deve ser justificada em nenhum sentido moral — é por isso que você deve aprender a controlar sua raiva."

Ou, “Na verdade, não existe nada como raiva — é uma ilusão, uma citação sem referente, como uma miragem de água no deserto, ou procurar na garagem por um dragão e não encontrar nenhum.”

Essas são pílulas difíceis de engolir (não que você deva engoli-las) e, portanto, é muito mais fácil professá-las do que acreditar nelas.

Acho que é isso que os não-reducionistas/não-materialistas pensam que estão criticando quando criticam o materialismo redutivo.

Mas o materialismo não é tão fácil. Não é tão barato quanto dizer, “A raiva é feita de átomos — pronto, agora terminei.” Isso não explicaria como ir de bolas de bilhar a bater. Você precisa dos insights específicos da computação, consequencialismo e árvores de busca antes de poder começar a fechar o abismo explicativo.

Tudo isso foi um exemplo relativamente fácil pelos padrões modernos, porque me restringi a falar sobre comportamentos raivosos. Falar sobre saídas não exige que você aprecie como um algoritmo se sente por dentro (cruzar um abismo de primeira pessoa/terceira pessoa) ou dissolver uma pergunta errada (desembaraçar lugares onde o interior de sua própria mente corre enviesado em relação à realidade).

Ir de substâncias materiais que dobram e quebram, queimam e caem, empurram e se chocam, para o comportamento raivoso, é apenas um problema de prática pelos padrões da filosofia moderna. Mas é um problema de prática importante. Só pode ser totalmente apreciado se você perceber o quão difícil teria sido resolvê-lo antes da invenção da escrita. Houve uma vez um abismo explicativo aqui — embora possa não parecer assim em retrospecto, agora que ele foi transposto por gerações.

Abismos explicativos podem ser cruzados, se você aceitar ajuda da ciência e não confiar na visão do interior de sua própria mente.

216 - Calor versus movimento

Depois do último ensaio, ocorreu-me haver um exemplo muito mais simples de reducionismo pulando um abismo de aparente diferença de espécie: a redução do calor ao movimento.

Hoje, a equivalência de calor e movimento pode parecer óbvia em retrospectiva - todo mundo diz que "calor é movimento", portanto, não pode ser uma crença "estranha".

Mas houve um tempo em que a teoria cinética do calor era uma hipótese científica altamente controversa, contrastando com a crença em um fluido calórico que fluía de objetos quentes para objetos frios. Ainda antes, a principal teoria do calor era "Flogisto!"

Suponha que você tivesse estudado separadamente a teoria cinética e a teoria calórica. Agora você sabe algo sobre cinética: colisões, rebatidas elásticas, momento, energia cinética, gravidade, inércia, trajetórias livres. Separadamente, você sabe algo sobre calor: temperaturas, pressões, combustão, fluxos de calor, motores, fusão, vaporização.

Esse estado de conhecimento não é apenas plausível, é o estado de conhecimento possuído, por exemplo, por Sadi Carnot, que, trabalhando estritamente na teoria calórica do calor, desenvolveu o princípio do ciclo de Carnot—um motor térmico de eficiência máxima, cuja existência implica a Segunda Lei da Termodinâmica. Isso em 1824, quando a cinética era uma ciência altamente desenvolvida.

Suponha, como Carnot, que você saiba muito sobre cinética e muito sobre calor, como entidades separadas. Entidades separadas de conhecimento, isto é: seu cérebro tem cestas de arquivamento separadas para crenças sobre cinética e crenças sobre calor. Mas, por dentro, esse estado de conhecimento parece viver em um mundo de coisas em movimento e coisas quentes, um mundo onde movimento e calor são propriedades independentes da matéria.

Agora, um Físico do Futuro aparece e lhe diz: "Onde há calor, há movimento, e vice-versa. É por isso, por exemplo, que esfregar coisas juntas as faz ficarem mais quentes."

Existem (pelo menos) duas interpretações possíveis que você poderia atribuir a essa afirmação, "Onde há calor, há movimento, e vice-versa."

Primeiro, você poderia supor que calor e movimento existem separadamente—que a teoria calórica está correta—mas que entre as leis físicas do nosso universo existe uma "lei de ligação" que afirma que, onde objetos estão se movendo rapidamente, o calórico surgirá. E, inversamente, outra lei de ligação diz que o calórico pode exercer pressão sobre as coisas e fazê-las se mover, e é por isso que um gás mais quente exerce mais pressão sobre seu recipiente (assim, um motor a vapor pode usar vapor para mover um pistão).

Segundo, você poderia supor que calor e movimento são, de alguma forma ainda misteriosa, a mesma coisa.

"Bobagem," diz o Pensador 1, "as palavras 'calor' e 'movimento' têm dois significados diferentes; é por isso que temos duas palavras diferentes. Sabemos como determinar quando chamaremos um fenômeno observado de 'calor' - calor pode derreter coisas ou fazê-las pegar fogo. Sabemos como determinar quando diremos que um objeto está 'se movendo rapidamente'—ele muda de posição; e quando colide, pode se de-

formar ou se quebrar. Calor está relacionado à mudança de substância; movimento, à mudança de posição e forma. Dizer que essas duas palavras têm o mesmo significado é simplesmente confundir a si mesmo."

"Impossível," diz o Pensador 2. "Pode ser que, em nosso mundo, calor e movimento estejam associados por leis de ligação, de modo que é uma lei da física que o movimento cria calórico, e vice-versa. Mas posso facilmente imaginar um mundo onde esfregar coisas juntas não as torna mais quentes, e gases não exercem mais pressão a temperaturas mais altas. Como existem mundos possíveis onde calor e movimento não estão associados, eles devem ser propriedades diferentes—isso é verdade a priori."

O Pensador 1 está confundindo a citação e o referente: 2 + 2 = 4, mas "2+2" ≠ "4." A string "2+2" contém cinco caracteres (incluindo espaços) e a string "4" contém apenas um caractere. Se você digitar as duas strings em um interpretador Python, elas produzem a mesma saída, >>> 4. Então você não pode concluir, ao olhar para as strings "2 + 2" e "4", que só porque as strings são diferentes, elas devem ter significados diferentes em relação ao Interpretador Python.

As palavras "calor" e "energia cinética" podem ser ditas como "referindo-se" à mesma coisa, mesmo antes de sabermos como o calor se reduz ao movimento, no sentido de que ainda não sabemos qual é o referente, mas os referentes são de fato os mesmos. Você pode imaginar um Interpretador de Ciência Idealizado e Onisciente que daria a mesma saída quando digitássemos "calor" e "energia cinética" na linha de comando.

Falo sobre o Interpretador de Ciência para enfatizar que, para desreferenciar o ponteiro, você precisa sair da cognição. O resultado da desreferenciação é algo lá fora na realidade, não na mente de ninguém. Então você pode dizer "referente real" ou "referente verdadeiro", mas não pode avaliar as palavras localmente, de dentro da sua própria cabeça. Você não pode raciocinar usando o calor real como referente—se você pensasse usando o calor real, pensar em "um milhão de Kelvin" vaporizaria seu cérebro. Mas, ao formar uma crença sobre sua crença a respeito do calor, você pode falar sobre sua crença sobre o calor e dizer coisas como "É possível que minha crença sobre o calor não se assemelhe muito ao calor real." Você não pode realmente realizar essa comparação na sua própria mente, mas pode falar sobre isso.

Portanto, você pode dizer: "Minhas crenças sobre calor e movimento não são as mesmas crenças, mas é possível que o calor real e o movimento real sejam a mesma coisa." É como conseguir reconhecer que "a estrela da manhã" e "a estrela da noite" podem ser o mesmo planeta, enquanto também entende que você não pode determinar isso apenas examinando suas crenças - você precisa pegar o telescópio.

O erro do Pensador 2 segue semelhantemente. Um físico lhes disse: "Onde há calor, há movimento" e o Pensador 2 confundiu isso com uma declaração de lei física: A presença de calórico causa a existência de movimento. O que o físico realmente quer dizer é mais semelhante a uma regra inferencial: onde lhe dizem haver "calor", deduza a presença de "movimento".

Dessa projeção básica de um modelo multinível em uma realidade multinível segue outro erro distinto: a confusão de possibilidade conceitual com possibilidade lógica. Para Sadi Carnot, é concebível que possa haver outro mundo onde calor e movimento não estejam associados. Para Richard Feynman, armado com conhecimento específico de como derivar equações sobre calor a partir de equações sobre movimento, essa ideia não é apenas inconcebível, mas tão selvagemente inconsistente que faz a cabeça explodir.

Devo notar, em justiça aos filósofos, que existem filósofos que disseram essas coisas. Por exemplo, Hilary Putnam, escrevendo sobre o experimento mental da "Terra Gêmea": [1]

> Uma vez que descobrimos que a água (no mundo real) é H_2O, nada conta como um mundo possível onde a água não seja H_2O. Em particular, se uma declaração "logicamente possível" é aquela que se mantém em algum "mundo logicamente possível", não é logicamente possível que a água não seja H_2O.
>
> Por outro lado, podemos perfeitamente imaginar ter experiências que nos convenceriam (e que tornariam racional acreditar que) a água não é H_2O. Nesse sentido, é concebível que a água não seja H_2O.

É concebível, mas não é logicamente possível! A concebibilidade não é prova de possibilidade lógica[27].

Parece-me que “água” está sendo usado em dois sentidos diferentes nesses dois parágrafos - um em que a palavra “água” se refere ao que digitamos no Interpretador de Ciência, e outro em que “água” se refere ao que obtemos do Interpretador de Ciência quando digitamos “água” nele. No primeiro parágrafo, Hilary parece estar dizendo que, após fazermos alguns experimentos e descobrirmos que a água é H_2O, a água se redefine automaticamente para significar H_2O. Mas você poderia defender coerentemente uma posição diferente sobre se a palavra “água” agora significa “H_2O” ou “o que realmente está naquela garrafa ao meu lado”, desde que você use seus termos consistentemente.

Acredito que o que foi dito acima já foi dito também? De qualquer forma…

É bastante possível que haja apenas uma coisa lá fora no mundo, mas que ela tome formas suficientemente diferentes, e que você mesmo seja suficientemente ignorante da redução, de modo que parece viver em um mundo contendo duas coisas completamente diferentes. O conhecimento sobre esses dois fenômenos diferentes pode ser ensinado em duas classes diferentes, e estudado por dois campos acadêmicos diferentes, localizados em dois prédios diferentes da sua universidade.

Você precisa se colocar bastante para trás, em uma mentalidade historicamente realista, para lembrar como calor e movimento uma vez pareceram diferentes. Embora, dependendo de quanto você sabe hoje, pode não ser tão difícil assim, se você conseguir olhar além da pressão da convencionalidade (ou seja, “calor é movimento” é uma crença não estranha, “calor não é movimento” é uma crença estranha). Quero dizer, suponha que amanhã os físicos dessem um passo à frente e dissessem: “Nossas popularizações da ciência sempre contiveram uma mentira. Na verdade, o calor não tem nada a ver com movimento”. Você poderia provar que eles estavam errados?

Dizer “Talvez calor e movimento sejam a mesma coisa!” é fácil. A parte difícil é explicar como. É preciso muito conhecimento detalhado para chegar ao ponto em que não se pode mais conceber um mundo no qual os dois fenômenos seguem caminhos separados. A redução não é barata, e é por isso que ela custa tanto.

Ou talvez você possa dizer: “O reducionismo é fácil, a redução é difícil”. Mas acho que ser reducionista ajuda um pouco, quando chega a hora de procurar uma redução.

Referências

[1] Hilary Putnam, “The Meaning of Meaning,” in The Twin Earth Chronicles, ed. Andrew Pessin and Sanford Goldberg (M. E. Sharpe, Inc., 1996), 3–52.

27 NT. Texto original em inglês. *Once we have discovered that water (in the actual world) is H_2O, nothing counts as a possible world in which water isn't H_2O. In particular, if a "logically possible" statement is one that holds in some "logically possible world," it isn't logically possible that water isn't H_2O.On the other hand, we can perfectly well imagine having experiences that would convince us (and that would make it rational to believe that) water isn't H_2O. In that sense, it is conceivable that water isn't H_2O. It is conceivable but it isn't logically possible! Conceivability is no proof of logical possibility.*

217 - Descoberta Revolucionária! O Cérebro é Feito de Neurônios!

Em uma descoberta surpreendente, uma equipe multinacional de cientistas liderada pelo ganhador do Prêmio Nobel Santiago Ramón y Cajal anunciou que o cérebro é composto por uma rede ridiculamente complexa de minúsculas células conectadas entre si por fios e ramificações infinitesimais.

A equipe multinacional - que também inclui o famoso técnico Antonie van Leeuwenhoek, e possivelmente Imhotep, promovido a deus egípcio da medicina - emitiu esta declaração:

"A presente descoberta culmina anos de pesquisa indicando que a coisa mole e convoluta dentro de nossos crânios é ainda mais complicada do que parece. Graças à aplicação por Cajal de uma nova técnica de coloração inventada por Camillo Golgi, aprendemos que essa estrutura não é uma rede contínua como os vasos sanguíneos do corpo, mas, na verdade, é composta por muitas células minúsculas, ou 'neurônios', conectadas umas às outras por filamentos ainda mais minúsculos.

Outras evidências extensas, começando com o pesquisador médico grego Alcmaeon e continuando com a pesquisa de Paul Broca sobre déficits de fala, indicam que o cérebro é a sede da razão.

Nemesius, o bispo de Emesia, argumentou anteriormente que o tecido cerebral é muito terreno para atuar como intermediário entre o corpo e a alma, e, portanto, as faculdades mentais estão localizadas nos ventrículos do cérebro. No entanto, se isso estiver correto, não há razão para que este órgão tenha uma composição interna imensamente complicada.

Charles Babbage sugeriu independentemente que muitos pequenos dispositivos mecânicos poderiam ser reunidos em um 'Motor Analítico', capaz de realizar atividades, como aritmética, que são amplamente consideradas como exigindo pensamento. O trabalho de Luigi Galvani e Hermann von Helmholtz sugere que as atividades dos neurônios são de natureza eletroquímica, em vez de pressões mecânicas como se acreditava anteriormente. No entanto, acreditamos que uma analogia com o 'Motor Analítico' de Babbage sugere que uma rede vastamente complicada de neurônios poderia igualmente exibir propriedades de pensamento.

Encontramos um sistema material enormemente complicado localizado onde a mente deveria estar. As implicações são chocantes e devem ser enfrentadas. Acreditamos que a presente pesquisa oferece forte evidência experimental de que Benedictus Spinoza estava certo e René Descartes errado: mente e corpo são de uma única substância.

Em combinação com o trabalho de Charles Darwin, mostrando como um órgão tão complicado poderia, em princípio, ter surgido como resultado de processos não inteligentes em si, a maior parte da evidência científica agora parece indicar que a inteligência não é ontologicamente fundamental e tem uma origem estendida no tempo. Isso pesa fortemente contra teorias que atribuem às entidades mentais um status ontologicamente fundamental ou causalmente primordial, incluindo todas as religiões já inventadas.

Ainda há muito trabalho a ser feito na descoberta das identidades específicas entre as interações eletroquímicas entre os neurônios e os pensamentos. No entanto, acreditamos que nossa descoberta oferece a promessa, embora ainda não a realização, de um relato científico completo do pensamento. O problema pode agora ser declarado, se não resolvido, então solucionável."

Lamentamos que Cajal e a maioria dos outros pesquisadores envolvidos no Projeto não estejam mais disponíveis para comentar.

218 - Quando o antropomorfismo se tornou estúpido

Acontece que a maioria das coisas no universo não tem mente.

Essa afirmação teria provocado incredulidade em muitas culturas antigas. "Animismo" é o termo usual. Eles pensavam que árvores, rochas, riachos e colinas tinham espíritos porque, ora, por que não?

Quer dizer, aqueles pedaços de carne conhecidos como "humanos" contêm pensamentos, então por que os pedaços de madeira conhecidos como "árvores" não deveriam conter?

Meus músculos se movem à minha vontade, e a água flui através de um rio. Quem pode dizer que o rio não tem vontade de mover a água? O rio transborda suas margens e inunda o local de reunião da minha tribo — por que não pensar que o rio estava com raiva, já que moveu suas partes para nos ferir? É o que pensaríamos quando o punho de alguém atingisse nosso nariz.

Não há razão óbvia — nenhuma razão óbvia para um caçador-coletor — porque isso não pode ser assim. Só parece um erro estúpido se você confundir estranheza com estupidez. Naturalmente, a crença de que os rios têm espíritos animadores parece "estranha" para nós, já que não é uma crença de nossa tribo. Mas não há nada obviamente estúpido em pensar que grandes massas de água em movimento têm espíritos, assim como nossas próprias massas de carne em movimento.

Se a ideia fosse obviamente estúpida, ninguém teria acreditado nela. Assim como, por muito tempo, ninguém acreditou na ideia obviamente estúpida de que a Terra se move enquanto parece imóvel.

É óbvio que as árvores não podem pensar? As árvores, não nos esqueçamos, são de fato nossas primas distantes. Volte longe o suficiente, e você tem um ancestral comum com sua samambaia. Se pedaços de carne podem pensar, por que não pedaços de madeira?

Para ser óbvio que a madeira não pensa, você precisa pertencer a uma cultura com microscópios. Não apenas quaisquer microscópios, mas microscópios realmente bons.

Aristóteles pensava que o cérebro era um órgão para resfriar o sangue. (É uma boa coisa que o que acreditamos sobre nossos cérebros tenha muito pouco efeito em sua operação real.)

Os egípcios jogavam fora o cérebro durante o processo de mumificação.

Alcméon de Crotona, um pitagórico do século V AEC, apontou o cérebro como a sede da inteligência, porque havia traçado o nervo óptico do olho até o cérebro. Ainda assim, com a quantidade de evidências que ele tinha, era apenas um palpite.

Quando o papel central do cérebro deixou de ser um palpite? Não conheço história o suficiente para responder a essa pergunta, e provavelmente não houve uma linha divisória nítida. Talvez possamos situá-lo no ponto em que alguém traçou a anatomia dos nervos e descobriu que cortar uma conexão nervosa com o cérebro bloqueava o movimento e a sensação?

Mesmo assim, isso é apenas um espírito misterioso movendo-se pelos nervos. Quem pode dizer que a madeira e a água, mesmo que não tenham os pequenos fios encontrados na anatomia humana, não possam carregar o mesmo espírito misterioso por meios diferentes?

Passei algum tempo online tentando rastrear o momento exato em que alguém notou a estrutura interna vastamente emaranhada dos neurônios do cérebro e disse: “Ei, aposto que todo esse emaranhado gigante está fazendo processamento complexo de informações!” Não tive muito sucesso. (Não foi Camillo Golgi — o emaranhado dos circuitos era conhecido antes de Golgi.) Talvez nunca tenha havido um momento decisivo ali também.

Mas a descoberta desse emaranhado, e a teoria da seleção natural de Charles Darwin, e a noção de cognição como computação, é onde eu colocaria o início gradual da descida do antropomorfismo para ser obviamente errado.

É o ponto em que você pode olhar para uma árvore e dizer: “Não vejo nada na biologia da árvore que esteja fazendo processamento complexo de informações. Nem vejo isso no comportamento, e se estiver escondido de uma forma que não afeta o comportamento da árvore, como surgiria uma pressão seletiva para tal processamento complexo de informações?”

É o ponto em que você pode olhar para um rio e dizer: “A água não contém padrões que se replicam com hereditariedade distante e variação substancial sujeita a seleção iterativa, então como um rio chegaria a ter qualquer padrão tão complexo e funcionalmente otimizado quanto um cérebro?”

É o ponto em que você pode olhar para um átomo e dizer: “A raiva pode parecer simples, mas não é, e não há espaço para ela caber em algo tão simples quanto um átomo - a menos que existam universos inteiros de sub-partículas dentro dos quarks; e mesmo assim, como nunca vimos nenhum sinal de raiva atômica, ela não teria nenhum efeito sobre os fenômenos de alto nível que conhecemos.”

É o ponto em que você pode olhar para um filhote de cachorro e dizer: “Os pais do filhote podem empurrá-lo para o chão quando ele faz algo errado, mas isso não significa que o filhote esteja fazendo raciocínio moral. Nossas teorias atuais de psicologia evolutiva afirmam que o raciocínio moral surgiu como uma resposta a desafios sociais mais complexos do que isso — em sua forma humana completa, nossas adaptações morais são o resultado de pressões seletivas sobre argumentos linguísticos acerca da política tribal.”

É o ponto em que você pode olhar para uma rocha e dizer: “Isso carece até mesmo das simples árvores de busca incorporadas em um programa de xadrez — de onde viria a intenção de querer rolar morro abaixo, como Aristóteles pensou uma vez?”

Está escrito:

Zhuangzi e Huizi estavam passeando ao longo da barragem da Cachoeira Hao quando Zhuangzi disse: “Veja como os peixinhos saem e nadam por onde querem! Isso é o que os peixes realmente gostam!”

Huizi disse: “Você não é um peixe — como você sabe do que os peixes gostam?”

Zhuangzi disse: “Você não é eu, então como você sabe que eu não sei do que os peixes gostam?”

Agora nós sabemos.

219 - A priori

A racionalidade tradicional é formulada como regras sociais, com violações interpretáveis como trapaça: se você quebra as regras e ninguém mais o faz, você é o primeiro a desertar—tornando-se uma pessoa muito ruim. Para os bayesianos, o cérebro é um motor de precisão: se você viola as leis da racionalidade, o motor não funciona, e isso é igualmente verdadeiro, independentemente de qualquer outra pessoa quebrar as regras ou não.

Considere o problema da Navalha de Ocam, conforme enfrentado pelos filósofos tradicionais. Se duas hipóteses se ajustam igualmente bem às mesmas observações, por que acreditar que a mais simples é mais provável de ser verdadeira? Você poderia argumentar que a Navalha de Ocam funcionou no passado e, portanto, é provável continuar a funcionar no futuro. Mas isso, por si só, apela a uma previsão da Navalha de Ocam. "A Navalha de Ocam funciona até 8 de outubro de 2027 e depois para de funcionar" é mais complexa, mas se ajusta igualmente bem às evidências observadas.

Você poderia argumentar que a Navalha de Ocam é uma distribuição razoável de probabilidades a priori. Mas o que é uma distribuição "razoável"? Por que não rotular "razoável" uma distribuição a priori muito complicada, que faz a Navalha de Ocam funcionar em todos os testes observados até agora, mas gera exceções em casos futuros?

De fato, parece não haver como *justificar* a Navalha de Ocam exceto *apelando* à própria Navalha de Ocam, tornando este *argumento* improvável de *convencer* qualquer juiz que não *aceite* já a Navalha de Ocam. (O que há de especial sobre as palavras que destaquei?)

Se você é um filósofo cujo trabalho diário é escrever artigos, criticar os artigos de outras pessoas e responder às críticas dos seus próprios artigos, então você pode olhar para a Navalha de Ocam e dar de ombros. Aqui está o fim de justificar, argumentar e convencer. Você decide dar uma trégua na escrita de artigos; se seus colegas filósofos não exigirem justificação para suas crenças inquestionáveis, você não exigirá justificação para as deles. E como símbolo do seu tratado, sua bandeira branca, você usa a frase "verdade a priori."

Mas para um bayesiano, nesta era de ciência cognitiva, biologia evolutiva e Inteligência Artificial, dizer "a priori" não explica por que o motor do cérebro funciona. Se o cérebro tem uma incrível "fábrica de verdades a priori" que funciona para produzir crenças precisas, isso faz você se perguntar por que um caçador-coletor sedento não pode usar a "fábrica de verdades a priori" para localizar água potável. Faz você se perguntar por que os olhos evoluíram em primeiro lugar, se existem maneiras de produzir crenças precisas sem olhar para as coisas.

James R. Newman disse: *"O fato de uma maçã somada a uma maçã resultar invariavelmente em duas maçãs ajuda no ensino da aritmética, mas não tem nenhuma relação com a verdade da proposição de que 1 + 1 = 2."* A Enciclopédia de Filosofia da Internet define proposições "a priori" como aquelas conhecíveis independentemente da experiência. A Wikipedia cita Hume: Relações de ideias são "descobertas pela mera operação do pensamento, sem dependência do que existe em qualquer lugar no universo." Você pode ver que 1 + 1 = 2 apenas pensando sobre isso, sem olhar para maçãs.

Mas nesta era da neurologia, deve-se estar ciente de que pensamentos existem no universo; eles são idênticos à operação dos cérebros. Cérebros materiais, reais no universo, compostos de quarks em uma física matemática unificada cujas leis não traçam uma fronteira entre o interior e o exterior do seu crânio.

Quando você soma 1 + 1 e obtém 2 pensando, esses pensamentos são eles próprios incorporados em flashes de padrões neurais. Em princípio, poderíamos observar, experimentalmente, os mesmos eventos materiais exatos enquanto ocorriam no cérebro de outra pessoa. Isso exigiria alguns avanços na neurobiologia computacional e na interface cérebro-computador, mas, em princípio, poderia ser feito. Você poderia ver o motor de outra pessoa operando materialmente, por meio de cadeias materiais de causa e efeito, para calcular por "pensamento puro" que 1 + 1 = 2. Como observar esse padrão no cérebro de outra pessoa é diferente, como forma de conhecimento, de observar seu próprio cérebro fazendo a mesma coisa? Quando o "pensamento puro" lhe diz que 1 + 1 = 2, "independentemente de qualquer experiência ou observação," você está, de fato, observando seu próprio cérebro como evidência.

Se isso parece contraintuitivo, tente ver mentes/cérebros como motores—um motor que colide o padrão neural para 1 e o padrão neural para 1 e obtém o padrão neural para 2. Se esse motor funciona, então ele deve ter a mesma saída se observar (com olhos e retina) um motor cerebral semelhante realizando uma colisão semelhante, e copiar para si mesmo o padrão resultante. Em outras palavras, para cada forma de conhecimento a priori obtida por "pensamento puro," você está aprendendo a mesma coisa que aprenderia se visse um motor cerebral externo realizando os mesmos flashes puros de ativação neural. Os motores são equivalentes, as saídas finais são equivalentes, os entrelaçamentos de crença são os mesmos.

Não há nada que você possa saber "a priori," que você não poderia saber com igual validade observando a liberação química de neurotransmissores dentro de algum cérebro externo. O que você acha que é, caro leitor?

É por isso que você pode prever o resultado de somar 1 maçã e 1 maçã imaginando-o primeiro em sua mente, ou digitar "3 x 4" em uma calculadora para prever o resultado de imaginar 4 filas com 3 maçãs por fila. Você e a maçã existem em um processo físico unificado sem fronteiras, e uma parte pode ecoar a outra.

Os tipos de flashes neurais que os filósofos rotulam como "crenças a priori" são arbitrários? Muitos algoritmos de IA funcionam melhor com "regularização" que faz o espaço das soluções tender em direção a soluções mais simples. Mas os algoritmos regularizados são eles próprios mais complexos; eles contêm uma linha extra de código (ou 1.000 linhas extras) em comparação com algoritmos não regularizados. O cérebro humano é tendencioso para a simplicidade, e pensamos de maneira mais eficiente assim. Se você pressionar o botão Ignorar neste ponto, você fica com um cérebro complexo que existe sem motivo e funciona sem motivo. Então, não tente me dizer que crenças "a priori" são arbitrárias, porque elas certamente não são geradas por números aleatórios. (O que o adjetivo "arbitrário" significa, afinal?)

Você não pode justificar chamar uma proposição de "a priori" apontando que outros filósofos estão tendo problemas para justificar suas proposições. Se um filósofo falha em explicar algo, esse fato não pode fornecer eletricidade a uma geladeira, nem atuar como uma fábrica mágica para crenças precisas. Não há trégua, não há bandeira branca, até você entender por que o motor funciona.

Se você limpar sua mente de justificações, de argumentos, então parece óbvio por que a Navalha de Ocam funciona na prática: vivemos em um mundo simples, um universo de baixa entropia em que há explicações curtas a serem encontradas. "Mas," você chora, "por que o próprio universo é ordenado?" Isso eu não sei, mas é o que vejo como o próximo mistério a ser explicado. Isso não é a mesma coisa que "Como eu argumento a Navalha de Ocam para um debatedor hipotético que ainda não a aceitou?"

Talvez você não possa argumentar nada para um debatedor hipotético que não aceitou a Navalha de Ocam, assim como você não pode argumentar nada para uma pedra. Uma mente precisa de uma certa quantidade de estrutura dinâmica para ser um aceitador de argumentos. Se uma mente não implementa Modus Ponens, ela pode aceitar "A" e "A → B" o dia todo sem nunca produzir "B." Como você justifica Modus Ponens para uma mente que não o aceitou? Como você argumenta com uma pedra para se tornar uma mente?

Os cérebros evoluíram de matéria não cerebral pela seleção natural; eles não foram justificados para existirem, argumentando com um estudante de filosofia ideal de vazio perfeito. Isso não torna nossos julgamentos sem sentido. Um motor cerebral pode funcionar corretamente, produzindo crenças precisas, mesmo se ele foi meramente construído - por mãos humanas ou pressões de seleção estocástica cumulativa - em vez de argumentado para existir. Mas para estar satisfeito com essa resposta, é preciso ver a racionalidade em termos de motores, e não de argumentos.

220 - Referência redutiva

A tese reducionista (como eu a formulo) é que as mentes humanas, por razões de eficiência, usam um mapa de múltiplos níveis no qual pensamos separadamente em coisas como "átomos" e "quarks", "mãos" e "dedos", ou "calor" e "energia cinética". A realidade em si, por outro lado, é de nível único, no sentido de que não parece conter átomos como entidades separadas, adicionais e causalmente eficazes além e acima dos quarks.

Sadi Carnot desenvolveu o (precursor da) Segunda Lei da Termodinâmica baseado na teoria calórica do calor, que considerava o calor como um fluido que passava de corpos quentes para corpos frios, produzido pelo fogo e causando a expansão dos gases. Os efeitos do calor eram estudados separadamente da ciência da cinética, muito antes de serem integrados. Se você está tentando projetar um motor a vapor, os efeitos de todas essas pequenas vibrações e colisões que chamamos de "calor" podem ser resumidos em uma descrição muito mais simples do que a mecânica quântica completa dos quarks. Os humanos calculam eficientemente, pensando apenas nos efeitos significativos sobre as quantidades relevantes para o objetivo.

A realidade, contudo, parece empregar a mecânica quântica completa dos quarks. Certa vez, conheci um sujeito que acreditava que, ao utilizar a Relatividade Geral para resolver um problema de baixa velocidade, como um projétil de artilharia, a Relatividade Geral forneceria uma resposta incorreta experimentalmente – não apenas uma resposta lenta, mas uma resposta experimentalmente incorreta - porque em baixas velocidades, os projéteis de artilharia são governados pela mecânica newtoniana, não pela Relatividade Geral. É exatamente assim que a física não funciona. A realidade parece continuar processando a Relatividade Geral, mesmo quando só faz diferença na décima quarta casa decimal, o que um humano consideraria um enorme desperdício de poder computacional. A física faz isso com força bruta. Ninguém nunca pegou a física simplificando seus cálculos - ou se alguém a pegou, os Senhores da Matrix apagaram a memória depois.

Nosso mapa, então, é muito diferente do território; nossos mapas são de vários níveis, o território é de nível único. Uma vez que a representação é tão incrivelmente diferente do referente, em que sentido uma crença como "estou usando meias" pode ser chamada de verdadeira, quando na realidade em si, existem apenas quarks?

Caso você tenha esquecido o que a palavra "verdadeiro" significa, a definição clássica foi dada por Alfred Tarski:

A afirmação "a neve é branca" é verdadeira se e somente se a neve for branca.

Caso você tenha esquecido qual é a diferença entre a afirmação "Eu acredito que 'a neve é branca'" e "'A neve é branca' é verdade", veja Qualitativamente Confuso. A verdade não pode ser avaliada apenas olhando dentro de sua própria cabeça - se você quiser saber, por exemplo, se "a estrela da manhã = a estrela da tarde", você precisa de um telescópio; não basta apenas olhar para as crenças em si.

Este é o ponto que os pós-modernistas não entendem quando gritam: "Mas como você sabe que suas crenças são verdadeiras?" Quando você faz um experimento, você realmente está saindo de sua própria cabeça. Você está se envolvendo em uma interação complexa cujo resultado é causalmente determinado pela coisa sobre a qual você está raciocinando, não apenas suas crenças sobre ela. Certa vez, defini "realidade" da seguinte forma:

> Mesmo quando tenho uma hipótese simples, fortemente apoiada por todas as evidências que conheço, às vezes ainda fico surpreso. Então, preciso de nomes diferentes para as coisas que determinam minhas previsões e a coisa que determina meus resultados experimentais. Chamo as primeiras coisas de "crença" e a última coisa de "realidade".

A interpretação do seu experimento ainda depende de suas crenças anteriores. Não falarei, por enquanto, sobre de onde vêm as crenças anteriores, porque esse não é o assunto deste ensaio. Meu ponto é que a verdade se refere a uma comparação ideal entre uma crença e a realidade. Como entendemos que os planetas são distintos das crenças sobre planetas, podemos projetar um experimento para testar se a crença "a estrela da manhã e a estrela da tarde são o mesmo planeta" é verdadeira. Este experimento envolverá telescópios, não apenas introspecção, porque entendemos que "verdade" envolve comparar uma crença interna a um fato externo; então usamos um instrumento, o telescópio, cujo comportamento percebido acreditamos depender do fato externo do planeta.

Acreditar que o telescópio nos ajuda a avaliar a "verdade" de "estrela da manhã = estrela da tarde" depende de nossas crenças anteriores sobre a interação do telescópio com o planeta. Novamente, não abordarei isso neste ensaio em particular, exceto para citar uma das minhas falas favoritas de Raymond Smullyan:

"Se o leitor mais sofisticado se opõe a esta afirmação por ser uma mera tautologia, então, por favor, pelo menos dê crédito à afirmação por não ser inconsistente". Da mesma forma, não vejo o uso de um telescópio como lógica circular, mas como coerência reflexiva; para cada forma sistemática de chegar à verdade, deve haver uma explicação racional de como ela funciona.

A questão em discussão é o que significa "a neve é branca" ser verdade, quando, na realidade, existem apenas quarks.

Há um certo padrão de conexões neurais que compõem suas crenças sobre "neve" e "brancura" - acreditamos nisso, mas não sabemos e não podemos visualizar concretamente as conexões neurais reais. Que estão, por sua vez, incorporadas em um padrão de quarks ainda menos conhecido. Lá fora, no mundo, existem moléculas de água cuja temperatura é baixa o suficiente para que se organizem em padrões repetidos em mosaico; elas não se parecem em nada com os emaranhados de neurônios. Em que sentido, comparando um padrão (sempre flutuante) de quarks com o outro, a crença "a neve é branca" é verdadeira?

Obviamente, nem eu, nem ninguém mais pode oferecer uma Função Ideal de Comparação de Quarks que aceite uma descrição ao nível de quark de uma crença neuralmente incorporada (incluindo o cérebro circundante) e uma descrição ao nível de quark de um floco de neve (e as leis da ótica circundantes), e produza "verdadeiro" ou "falso" sobre "a neve é branca". E quem disse que o nível fundamental é realmente sobre campos de partículas?

Por outro lado, descartar todas as crenças porque elas não estão escritas como especificações gigantescas e incontroláveis sobre quarks que nem podemos ver… não parece uma ideia muito prudente. Não é a melhor maneira de otimizar nossos objetivos.

Parece-me que uma palavra como "neve" ou "branco" pode ser vista como uma nota promissória - não uma especificação conhecida de quais configurações físicas de quarks contam como "neve", mas, mesmo assim, existem coisas chamadas de neve e coisas que não são, e mesmo que haja alguns itens que possam ser confundidos com neve (como neve de plástico), um Intérprete Científico Onisciente Ideal conseguiria identificar o centro da neve e redefinir o limite para uma definição mais simples.

Em um universo de camada única cujo nível inferior é desconhecido, ou incerto, ou simplesmente grande demais para falar sobre, os conceitos em uma mente de múltiplas camadas podem ser considerados como representando uma espécie de nota promissória - não sabemos a que eles correspondem, lá fora. Mas parece-nos que podemos distinguir casos positivos de negativos, de uma forma preditivamente produtiva, então pensamos - talvez em um sentido totalmente geral - que há alguma diferença de quarks, alguma diferença de configurações no nível fundamental, que explica as diferenças que alimentam nossos sentidos e, finalmente, resultam em dizermos "neve" ou "não neve".

Eu vejo essa coisa branca, e é a mesma em várias ocasiões, então levanto a hipótese de uma causa

latente estável no ambiente - dou a ela o nome de “neve”; “neve” é então uma nota promissória referindo-se a um limite simples acreditado que poderia ser traçado em torno das causas invisíveis da minha experiência.

O experimento mental “Terra Gêmea” de Hilary Putnam (onde a água não é H_2O, mas alguma outra substância estranha denotada XYZ, caso contrário se comportando como água) e o debate filosófico subsequente, ajuda a destacar essa questão. “Neve” não tem uma definição lógica conhecida por nós - é mais como um ponteiro determinado empiricamente para uma definição lógica. Isso é verdade mesmo se você acredita que a neve é formada por cristais de gelo, que são moléculas de água em mosaico a baixa temperatura. As moléculas de água são feitas de quarks. E se os quarks se revelarem feitos de outra coisa? O que é um floco de neve, então? Você não sabe - mas ainda é um floco de neve, não um hidrante.

E, claro, esses mesmos parágrafos que acabei de escrever estão da mesma forma muito acima do nível dos quarks. “Sentir coisas brancas, categorizá-las visualmente e pensar ‘neve’ ou ‘não neve’” - isso também está falando muito acima do nível dos quarks. Então, minhas meta-crenças também são notas promissórias, para coisas que um Intérprete Científico Onisciente Ideal poderia saber sobre quais configurações dos quarks (ou o que quer que seja) que compõem meu cérebro, correspondem a “acreditar que ‘a neve é branca’”.

Mas então, toda a compreensão que temos da realidade é composta de notas promissórias desse tipo. Portanto, em vez de chamá-la de circular, prefiro chamá-la de autoconsistente.

Isso pode ser um pouco enervante - manter um poleiro epistêmico precário, tanto em crenças de nível de objeto quanto em reflexão, muito acima de uma enorme realidade fundamental desconhecida subjacente, e esperar não cair.

Na reflexão, porém, é difícil ver como as coisas poderiam ser de outra maneira.

Então, no final do dia, a afirmação “a realidade não contém mãos como entidades causais fundamentais, adicionais e separadas, além dos quarks” não é a mesma afirmação que “mãos não existem” ou “eu não tenho mãos”. Não existem mãos fundamentais; as mãos são feitas de dedos, palma e polegar, que por sua vez são feitos de músculo e osso, até chegar aos campos de partículas elementares, que são as entidades causais fundamentais, até onde sabemos atualmente.

Isso não é o mesmo que dizer “não existem ‘mãos’”. Não é o mesmo que dizer “a palavra ‘mãos’ é uma nota promissória que nunca será paga, porque não há nenhum agrupamento empírico que corresponda a ela”; ou “a nota ‘mãos’ nunca será paga, porque é logicamente impossível reconciliar suas supostas características”; ou “a afirmação ‘humanos têm mãos’ se refere a um estado de coisas sensato, mas a realidade não está nesse estado”.

Apenas: Existem padrões que existem na realidade onde vemos “mãos”, e esses padrões têm algo em comum, mas não são fundamentais.

Se eu realmente não tivesse mãos - se a realidade de repente fizesse a transição para um estado que descreveríamos como “Eliezer não tem mãos” - a realidade logo depois corresponderia a um estado que descreveríamos como “Eliezer grita enquanto sangue jorra de seus tocos de pulso”.

E isso é verdade, mesmo que o parágrafo acima não tenha especificado nenhuma posição de quark.

A frase anterior é da mesma forma meta-verdadeira.

O mapa é multinível, o território é de nível único. Isso não significa que os níveis mais altos “não existem”, como procurar um dragão em sua garagem e não encontrar nada lá, ou como ver uma miragem no deserto e formar uma expectativa de água potável quando não há nada para beber. Os níveis mais altos do seu mapa não são falsos, sem referentes; eles têm referentes no nível único da física. Não é que as asas de um avião não existam - então o avião cairia do céu. As “asas de um avião” existem explicitamente no modelo multinível de um engenheiro de um avião, e as asas de um avião existem implicitamente na física quântica do avião real. A existência implícita não é o mesmo que inexistência. A descrição exata dessa implicitidade não é conhecida por nós - não é explicitamente representada em nosso mapa. Mas isso não impede que nosso

mapa funcione, ou até mesmo impede que seja verdadeiro.

Embora seja um pouco enervante contemplar que cada conceito e crença em seu cérebro, incluindo esses meta-conceitos sobre como seu cérebro funciona e por que você pode formar crenças precisas, estão empoleirados ordens e ordens de magnitude acima da realidade…

221 - Zumbis! Zumbis?

Seu “zumbi”, no uso filosófico do termo, é supostamente um ser exatamente igual a você em todos os aspectos - comportamento idêntico, fala idêntica, cérebro idêntico; cada átomo e quark na mesma posição, movendo-se conforme as mesmas leis causais de movimento - exceto que seu zumbi não é consciente.

Além disso, afirma-se que se zumbis são “possíveis” (um termo sobre o qual as batalhas continuam sendo travadas), então, apenas a partir de nosso conhecimento dessa “possibilidade”, podemos deduzir a priori que a consciência é extrafísica, em um sentido a ser descrito abaixo; o termo padrão para essa posição é “epifenomenalismo”.

(Para aqueles que não estão familiarizados com zumbis, enfatizo que isso não é um espantalho. Veja, por exemplo, a entrada da Stanford Encyclopedia of Philosophy sobre Zumbis. A “possibilidade” de zumbis é aceita por uma fração substancial, possivelmente a maioria, dos filósofos acadêmicos da consciência.)

Li em algum lugar: “Você não é aquele que fala seus pensamentos - você é aquele que ouve seus pensamentos.” Em hebraico, a palavra para a alma mais elevada, aquela que Deus soprou em Adão, é N’Shama - “o ouvinte”.

Se você concebe a “consciência” como uma escuta puramente passiva, então a noção de um zumbi inicialmente parece fácil de imaginar. É alguém que não tem o N’Shama, o ouvinte.

(Aviso: Ensaio muito longo de 6.600 palavras envolvendo David Chalmers à frente. Isso pode ser tomado como meu contra-exemplo demonstrativo à Parte II de Argumentando com Eliezer de Richard Chappell, na qual Richard me acusa de não me envolver com os argumentos complexos de filósofos reais.)

Quando você abre uma geladeira e descobre que o suco de laranja acabou, você pensa “Droga, estou sem suco de laranja”. O som dessas palavras é provavelmente representado em seu córtex auditivo, como se você tivesse ouvido outra pessoa dizer isso. (Por que penso isso? Porque falantes nativos de chinês podem se lembrar de sequências de dígitos mais longas do que falantes de inglês. Os dígitos chineses são todos monossilábicos, e os falantes de chinês podem se lembrar de cerca de dez dígitos, contra o famoso “sete mais ou menos dois” para falantes de inglês. Parece haver um loop de repetição de sons de volta para si, um limite de tamanho na memória de trabalho no córtex auditivo, o qual é genuinamente baseado em fonemas.)

Suponhamos que o exposto acima esteja correto; como um postulado, certamente não deve apresentar nenhum problema para os defensores dos zumbis. Mesmo que os humanos não sejam assim, parece fácil imaginar uma IA construída dessa maneira (e a imaginabilidade é o que o argumento zumbi trata). Não é apenas concebível em princípio, mas bem possível nas próximas décadas, que os cirurgiões coloquem uma rede de torneiras neurais sobre o córtex auditivo de alguém e leiam sua narrativa interna. (Pesquisadores já acessaram o núcleo geniculado lateral de um gato e reconstruíram entradas visuais reconhecíveis.)

Então, seu zumbi, sendo fisicamente idêntico a você até o último átomo, abrirá a geladeira e formará padrões corticais auditivos para os fonemas “Droga, estou sem suco de laranja”. Sobre este ponto, os epifenomenalistas concordariam de bom grado.

Mas, diz o epifenomenalista, no zumbi não há ninguém dentro para ouvir; o ouvinte interno está faltando. A narrativa interna é falada, mas não ouvida. Você não é aquele que fala seus pensamentos. Você

é aquele que os ouve.

Parece muito mais simples (eles diriam) fazer uma IA que imprima algum tipo de narrativa interna do que mostrar que um ouvinte interno a ouve.

O Argumento Zumbi é que se o Mundo Zumbi é possível - não necessariamente fisicamente possível em nosso universo, apenas "possível em teoria" ou "imaginável", ou algo nesse sentido - então a consciência deve ser extrafísica, algo além dos meros átomos. Por quê? Porque mesmo que você de alguma forma soubesse as posições de todos os átomos no universo, você ainda precisaria ser informado, como um fato separado e adicional, que as pessoas eram conscientes, que elas tinham ouvintes internos, que não estávamos no Mundo Zumbi, como parece possível.

O zumbi-ismo não é o mesmo que o dualismo. Descartes pensava haver uma substância corporal e um tipo totalmente diferente de substância mental, mas Descartes também pensava que a substância mental era um princípio causalmente ativo, interagindo com a substância corporal, controlando nossa fala e comportamento. Subtrair a substância mental do humano deixaria um zumbi tradicional, do tipo que se contorce e geme.

E embora a palavra hebraica para a alma mais íntima seja *N'Shama* (aquele que ouve), não me lembro de ter ouvido um rabino defendendo a possibilidade de zumbis. A maioria dos rabinos ficaria provavelmente horrorizada com a ideia de que a parte divina que Deus soprou em Adão não faz realmente nada.

O termo técnico para a crença de que a consciência está lá, mas não tem efeito no mundo físico, é epifenomenalismo.

Embora existam outros elementos no argumento zumbi (tratarei deles abaixo), acho que a intuição do ouvinte passivo é o que primeiro seduz as pessoas para o zumbi-ismo. Em particular, é o que seduz um público leigo ao zumbi-ismo. A noção central é simples e fácil de acessar: as luzes estão acesas, mas não há ninguém em casa.

Os filósofos estão apelando para a intuição do ouvinte passivo quando dizem "Claro que o mundo zumbi é imaginável; você sabe exatamente como seria".

Uma das grandes batalhas nas Guerras Zumbis é sobre o que, exatamente, significa dizer que zumbis são "possíveis". Os primeiros filósofos zumbi-istas (na década de 1970) apenas pensavam ser óbvio que os zumbis eram "possíveis" e não se preocuparam em definir a que tipo de possibilidade se referia.

Devido à minha leitura em lógica matemática, o que instantaneamente me vem à mente é a possibilidade lógica. Se você tem uma coleção de afirmações como {(A → B); (B → C); (C → ¬A)}, então a crença composta é logicamente possível se tiver um modelo - que, no caso simples acima, se reduz a encontrar uma atribuição de valor a {A;B;C} que torne todas as afirmações (A → B), (B → C) e (C → ¬A) verdadeiras. Neste caso, A = B = C = 0 funciona, assim como {A = 0;B = C = 1} ou {A = B = 0;C = 1}.

Algo parecerá possível - parecerá "conceitualmente possível" ou "imaginável" - se você puder considerar a coleção de afirmações sem ver uma contradição. Mas é, em geral, um problema muito difícil ver contradições ou encontrar um modelo específico completo! Se você se limitar a proposições booleanas simples da forma ((A ou B ou C) e (B ou ¬C ou D) e (D ou ¬A ou ¬C)...), conjunções de disjunções de três variáveis, então este é um problema muito famoso chamado 3-SAT, o qual é um dos primeiros problemas a ser comprovado NP--completo.

Portanto, só porque você não vê uma contradição no Mundo Zumbi à primeira vista, não significa que não haja contradição lá. É como não ver uma contradição na Hipótese de Riemann à primeira vista. Da possibilidade conceitual ("Não vejo problema") à possibilidade lógica, no sentido técnico completo, é um salto muito grande. É fácil torná-lo um salto NP-completo, e com teorias de primeira ordem você pode torná-lo arbitrariamente difícil de calcular, mesmo para perguntas finitas. E é a possibilidade lógica do Mundo Zumbi, não a possibilidade conceitual, que é necessária para supor que uma mente logicamente onisciente poderia saber as posições de todos os átomos no universo e ainda assim precisar ser informada como um fato adicional não implicado de que temos ouvintes internos.

Só porque você ainda não vê uma contradição, não é garantia de que você não verá uma contradição em mais trinta segundos. “Todos os números ímpares são primos. Prova: 3 é primo, 5 é primo, 7 é primo…”

Então, vamos ponderar o Argumento Zumbi um pouco mais: podemos pensar em um contra-exemplo para a afirmação “A consciência não tem impacto causal detectável por terceiros no mundo”?

Se você fechar os olhos e se concentrar em sua consciência interior, começará a formar pensamentos, em sua narrativa interna, que seguem as linhas de “Estou consciente” e “Minha consciência está separada dos meus pensamentos” e “Eu não sou aquele que fala meus pensamentos, mas aquele que os ouve” e “Meu fluxo de consciência não é minha consciência” e “Parece haver uma parte de mim que posso imaginar sendo eliminada sem mudar meu comportamento exterior”.

Você pode até dizer essas frases em voz alta, enquanto medita. Em princípio, alguém com um super-fMRI provavelmente poderia ler os fonemas do seu córtex auditivo; mas dizê-lo em voz alta remove todas as dúvidas sobre se você entrou nos reinos da testabilidade e das consequências físicas.

Isso certamente parece que o ouvinte interno está sendo pego no ato de ouvir por qualquer parte de você que escreve a narrativa interna e move seus lábios.

Imagine que uma misteriosa raça de alienígenas o visite e lhe deixe uma misteriosa caixa preta como presente. Você tenta cutucar e sondar a caixa preta, mas (até onde você pode dizer) nunca consegue obter uma reação. Você não consegue fazer a caixa preta produzir moedas de ouro ou responder a perguntas. Então você conclui que a caixa preta é causalmente inativa: “Para todo X, a caixa preta não faz X”. A caixa preta é um efeito, mas não uma causa; epifenomenal; sem potência causal. Em sua mente, você testa essa hipótese geral para ver se ela é verdadeira em alguns casos de teste, e parece verdadeira - “A caixa preta transforma chumbo em ouro? Não. A caixa preta ferve água? Não”.

Mas você pode ver a caixa preta; ela absorve luz e pesa na sua mão. Isso também faz parte da dança da causalidade. Se a caixa preta estivesse totalmente fora do universo causal, você não poderia vê-la; você não teria como saber que ela existe; você não poderia dizer “Obrigado pela caixa preta”. Você não pensou nesse contra-exemplo quando formulou a regra geral: “Todo X: Caixa preta não faz X”. Mas estava lá o tempo todo.

(Na verdade, os alienígenas deixaram outra caixa preta para você, esta puramente epifenomenal, e você não tem a menor ideia de que ela está lá em sua sala de estar. Essa foi a piada deles.)

Se você conseguir fechar seus olhos e sentir-se sentindo - se estiver consciente de si mesmo estando consciente, e pensar “Estou ciente de que estou consciente” - e disser em voz alta “Estou ciente de que estou consciente” - então sua consciência não está sem efeito em sua narrativa interna ou em seus lábios em movimento. Você pode perceber-se vendo, e sua narrativa interna reflete isso, assim como seus lábios se você decidir dizê-lo em voz alta.

Eu não vi o argumento acima escrito dessa maneira particular - “o ouvinte pego no ato de ouvir” - embora possa muito bem ter sido dito antes.

Mas é um ponto padrão - que os filósofos zumbi-istas aceitam! - que os filósofos do Mundo Zumbi, sendo átomo por átomo idênticos aos nossos próprios filósofos, escrevem artigos idênticos sobre a filosofia da consciência.

Neste ponto, o Mundo Zumbi deixa de ser uma consequência intuitiva da ideia de um ouvinte passivo.

Filósofos escrevendo artigos sobre consciência parecem ser pelo menos um efeito da consciência sobre o mundo. Você pode argumentar razões inteligentes pelas quais isso não é assim, mas você tem que ser inteligente.

Você intuitivamente suporia que se sua consciência interior desaparecesse, isso mudaria o mundo, pois sua narrativa interna não diria mais coisas como “Há um ouvinte misterioso dentro de mim”, porque o ouvinte misterioso teria desaparecido. Geralmente é logo depois que você concentra sua consciência em sua

consciência que sua narrativa interna diz "Estou ciente da minha consciência", o que sugere que se o primeiro evento nunca mais acontecesse, o segundo também não aconteceria. Você pode argumentar razões inteligentes pelas quais isso não é assim, mas você tem que ser inteligente.

É possível formar uma crença proposicional de que "A consciência não tem efeito" e, a princípio, não perceber nenhuma contradição se não compreender que falar sobre consciência é um efeito de estar consciente. Porém, uma vez que se vê a conexão entre a regra geral de que a consciência não tem efeito e a implicação específica de que a consciência não influencia como os filósofos escrevem artigos sobre consciência, o zumbi-ismo deixa de ser intuitivo e exige a postulação de coisas estranhas.

Uma coisa estranha que você pode postular é haver um Mestre Zumbi, um deus no Mundo Zumbi que secretamente assume o controle dos filósofos zumbis e os faz falar e escrever sobre consciência.

Um Mestre Zumbi não parece impossível. Os seres humanos muitas vezes não parecem tão coerentes ao falar sobre consciência. Pode não ser tão difícil falsificar seu discurso, para os padrões de, digamos, um amador humano falando em um bar. Talvez você pudesse pegar, como um corpus, mil amadores humanos tentando discutir a consciência; alimentá-los em uma IA não consciente, mas sofisticada, melhor do que os modelos de hoje, mas não auto-modificável; e obter um discurso sobre "consciência" que soasse tão sensato quanto a maioria dos humanos, ou seja, não muito.

Mas esse discurso sobre "consciência" não seria espontâneo. Não seria produzido dentro da IA. Seria uma imitação gravada de outra pessoa falando. Isso é apenas um *holodeck*[28], com uma IA central escrevendo o discurso dos personagens não-jogadores. Não é disso que se trata o Mundo Zumbi.

Por suposição, o Mundo Zumbi é idêntico ao nosso, átomo por átomo, exceto que os habitantes não têm consciência. Além disso, os átomos no Mundo Zumbi se movem sob as mesmas leis da física que em nosso próprio mundo. Se existem "leis de ligação" que governam quais configurações de átomos evocam a consciência, essas leis de ligação estão ausentes. Mas, por hipótese, a diferença não é experimentalmente detectável. Quando se trata de dizer se um quark ziguezagueia ou exerce uma força sobre os quarks próximos - qualquer coisa experimentalmente mensurável - as mesmas leis físicas governam.

O Mundo Zumbi não tem espaço para um Mestre Zumbi, porque um Mestre Zumbi tem que controlar os lábios do zumbi, e esse controle é, em princípio, experimentalmente detectável. O Mestre Zumbi move os lábios, portanto, tem consequências observáveis. Haveria um ponto em que um elétron faria um zag, em vez de um zig, porque o Mestre Zumbi assim o diz. (A menos que o Mestre Zumbi esteja realmente no mundo, como um padrão de quarks - mas então o Mundo Zumbi não é idêntico ao nosso, átomo por átomo, a menos que você pense que este mundo também contém um Mestre Zumbi.)

Quando um filósofo em nosso mundo digita "Eu acho que o Mundo Zumbi é possível", seus dedos tocam as teclas em sequência: z-u-m-b-i. Há uma cadeia de causalidade que pode ser rastreada a partir dessas teclas: músculos se contraindo, nervos disparando, comandos enviados pela medula espinhal, do córtex motor - e então para áreas menos compreendidas do cérebro, onde a narrativa interna do filósofo começou a falar sobre "consciência".

E o gêmeo zumbi do filósofo toca as mesmas teclas, pela mesma razão, causalmente falando. Não há causa na cadeia de explicação para o porquê de o filósofo escrever da maneira que ele escreve que não esteja também presente no gêmeo zumbi. O gêmeo zumbi também tem uma narrativa interna sobre "consciência", que um super-fMRI poderia ler do córtex auditivo. E quaisquer outros pensamentos, ou outras causas de qualquer tipo, que levaram a essa narrativa interna, são os mesmos em nosso próprio universo e no Mundo Zumbi.

Então você não pode dizer que o filósofo está escrevendo sobre consciência devido à consciência,

28 NT. **Holodeck**: um ambiente virtual fictício presente na franquia "Star Trek", capaz de criar simulações realistas por meio de hologramas e campos de força. Essa tecnologia imaginária permite vivenciar qualquer cenário ou situação, servindo tanto para treinamento quanto entretenimento. O conceito influenciou discussões sobre realidade virtual e suas possibilidades.

enquanto o gêmeo zumbi está escrevendo sobre consciência devido a um Mestre Zumbi ou chatbot de IA. Quando você rastreia a cadeia de causalidade por trás do teclado, até a narrativa interna ecoada no córtex auditivo, até a causa da narrativa, você deve encontrar a mesma explicação física em nosso mundo e no mundo zumbi.

Como o mais formidável defensor do zumbi-ismo, David Chalmers, escreve: [1]

> Pense no meu gêmeo zumbi no universo ao lado. Ele fala sobre experiência consciente o tempo todo - na verdade, ele parece obcecado por ela. Ele passa quantidades ridículas de tempo debruçado sobre um computador, escrevendo capítulo após capítulo sobre os mistérios da consciência. Ele frequentemente comenta sobre o prazer que obtém de certas qualia sensoriais, professando um amor particular por verdes e roxos profundos. Ele frequentemente entra em discussões com materialistas zumbis, argumentando que sua posição não pode fazer justiça às realidades da experiência consciente.
>
> E ainda assim ele não tem nenhuma experiência consciente! Em seu universo, os materialistas estão certos e ele está errado. A maioria de suas afirmações sobre a experiência consciente é totalmente falsa. Mas certamente há uma explicação física ou funcional de por que ele faz as afirmações que faz. Afinal, seu universo é totalmente governado por leis, e nenhum evento nele é milagroso, então deve haver alguma explicação para suas afirmações.
>
> … Qualquer explicação do comportamento do meu gêmeo contará igualmente como uma explicação do meu comportamento, já que os processos dentro de seu corpo são precisamente espelhados por aqueles dentro do meu. A explicação de suas afirmações obviamente não depende da existência da consciência, já que não há consciência em seu mundo. Segue-se que a explicação das minhas afirmações também é independente da existência da consciência.

Chalmers não está argumentando contra zumbis; essas são suas crenças reais!

> Essa situação paradoxal é ao mesmo tempo, deliciosa e perturbadora. Não é obviamente fatal para a posição não reducionista, mas é pelo menos algo com o qual precisamos lidar[29]…

Eu seriamente indicaria isso como a maior bala já mordida na história do tempo. E isso é um elogio indireto a David Chalmers: um mortal menor simplesmente não veria as implicações, ou se recusaria a enfrentá-las, ou racionalizaria uma razão para não ser assim.

Por que alguém morderia uma bala tão grande? Por que alguém postularia zumbis inconscientes que escrevem artigos sobre consciência exatamente pela mesma razão que nossos próprios filósofos genuinamente conscientes fazem?

Não devido à primeira intuição que escrevi, a intuição do ouvinte passivo. Essa intuição pode dizer que zumbis podem dirigir carros ou fazer matemática, ou até mesmo se apaixonar, mas não diz que zumbis escrevem artigos de filosofia sobre seus ouvintes passivos.

O argumento zumbi não se baseia apenas na intuição do ouvinte passivo. Se isso fosse tudo o que havia no argumento zumbi, ele já estaria morto, eu acho. A intuição de que o "ouvinte" pode ser eliminado

---

29 NT. Texto original em inglês. *Think of my zombie twin in the universe next door. He talks about conscious experience all the time—in fact, he seems obsessed by it. He spends ridiculous amounts of time hunched over a computer, writing chapter after chapter on the mysteries of consciousness. He often comments on the pleasure he gets from certain sensory qualia, professing a particular love for deep greens and purples. He frequently gets into arguments with zombie materialists, arguing that their position cannot do justice to the realities of conscious experience. And yet he has no conscious experience at all! In his universe, the materialists are right and he is wrong. Most of his claims about conscious experience are utterly false. But there is certainly a physical or functional explanation of why he makes the claims he makes. After all, his universe is fully law-governed, and no events therein are miraculous, so there must be some explanation of his claims. . . . Any explanation of my twin's behavior will equally count as an explanation of my behavior, as the processes inside his body are precisely mirrored by those inside mine. The explanation of his claims obviously does not depend on the existence of consciousness, as there is no consciousness in his world. It follows that the explanation of my claims is also independent of the existence of consciousness. This paradoxical situation is at once delightful and disturbing. It is not obviously fatal to the nonreductive position, but it is at least something that we need to come to grips with . . .*

sem efeito desapareceria assim que você percebesse que sua narrativa interna rotineiramente parece pegar o ouvinte no ato de ouvir.

Não, o impulso para morder essa bala vem de uma intuição totalmente diferente - a intuição de que não importa quantos átomos você some, não importa quantas massas e cargas elétricas interajam entre si, elas nunca necessariamente produzirão uma sensação subjetiva da misteriosa vermelhidão do vermelho. Pode ser um fato sobre nosso universo físico (diz Chalmers) que colocar tais e tais átomos em tal e tal posição evoca uma sensação de vermelhidão; mas se assim for, não é um fato necessário, é algo a ser explicado acima e além do movimento dos átomos.

Mas se você considerar a segunda intuição por si só, sem a intuição do ouvinte passivo, é difícil ver por que ela implica em zumbi-ismo. Talvez haja apenas um tipo diferente de coisa, além e adicional aos átomos, que não é causalmente passiva - uma alma que realmente faz coisas, uma alma que desempenha um papel causal real em por que escrevemos sobre "a misteriosa vermelhidão do vermelho". Retire a alma, e... bem, assumindo que você não apenas caia em coma, você certamente não escreverá mais artigos sobre consciência!

Esta é a posição tomada por Descartes e a maioria dos outros pensadores antigos: A alma é de um tipo diferente, mas interage com o corpo. A posição de Descartes é tecnicamente conhecida como dualismo de substância - há uma substância de pensamento, uma substância mental, e não é como os átomos; mas é causalmente potente, interativa e deixa uma marca visível em nosso universo.

Os zumbi-istas são dualistas de propriedade - eles não acreditam em uma alma separada; eles acreditam que a matéria em nosso universo tem propriedades adicionais além das físicas.

"Além do físico"? O que isso significa? Significa que as propriedades extras estão lá, mas não influenciam o movimento dos átomos, como as propriedades de carga elétrica ou massa. As propriedades extras não são experimentalmente detectáveis por terceiros; você sabe que está consciente, por dentro de suas propriedades extras, mas nenhum cientista pode detectar isso diretamente de fora.

Então, as propriedades adicionais estão lá, mas não são causalmente ativas. As propriedades extras não movem os átomos, e é por isso que elas não podem ser detectadas por terceiros.

E é por isso que podemos (supostamente) imaginar um universo como este, com todos os átomos nos mesmos lugares, mas as propriedades extras ausentes, para que tudo continue igual a antes, mas ninguém esteja consciente.

O Mundo Zumbi pode não ser fisicamente possível, dizem os zumbi-istas - porque é um fato que toda a matéria em nosso universo tem as propriedades extras, ou obedece às leis de ligação que evocam a consciência - mas o Mundo Zumbi é logicamente possível: as leis de ligação poderiam ter sido diferentes.

Mas, uma vez que você percebe que a concebibilidade não é o mesmo que a possibilidade lógica, e que o Mundo Zumbi nem é tão intuitivo, por que dizer que o Mundo Zumbi é logicamente possível?

Por que, oh! por que, dizer que as propriedades extras são epifenomenais e indetectáveis?

Podemos colocar este dilema de forma bem clara: Chalmers acredita haver algo chamado consciência, e essa consciência incorpora a verdadeira e indescritível substância da misteriosa vermelhidão do vermelho. Pode ser uma propriedade além da massa e da carga, mas está lá, e é a consciência. Agora, tendo dito isso, Chalmers especifica ainda que essa verdadeira substância da consciência é epifenomenal, sem potência causal - mas por que dizer isso?

Por que dizer que você poderia subtrair essa verdadeira substância da consciência e deixar todos os átomos no mesmo lugar fazendo as mesmas coisas? Se isso for verdade, precisamos de alguma explicação física separada para o porquê de Chalmers falar sobre "a misteriosa vermelhidão do vermelho". Ou seja, existe tanto uma misteriosa vermelhidão do vermelho, o qual é extrafísico, quanto uma razão inteiramente separada, na física, pela qual Chalmers fala sobre a "misteriosa vermelhidão do vermelho".

Chalmers confessa que essas duas coisas parecem que deveriam estar relacionadas, mas realmente, por que você precisa de ambas? Por que não escolher apenas uma ou a outra?

Uma vez que você postulou haver uma misteriosa vermelhidão do vermelho, por que não apenas dizer que ele interage com sua narrativa interna e faz você falar sobre a "misteriosa vermelhidão do vermelho"?

Descartes não está adotando a abordagem mais simples aqui? A abordagem estritamente mais simples?

Por que postular uma alma extra material e depois postular que a alma não tem efeito no mundo físico e depois postular um misterioso processo material desconhecido que faz sua narrativa interna falar sobre a experiência consciente?

Por que não postular a verdadeira substância da consciência que nenhuma quantidade de meros átomos mecânicos pode se somar, e então, já tendo ido tão longe, deixar essa verdadeira substância da consciência ter efeitos causais como fazer os filósofos falarem sobre consciência?

Eu não estou endossando a visão de Descartes. Mas pelo menos posso entender de onde Descartes está vindo. A consciência parece misteriosa, então você postula uma substância misteriosa da consciência. Tudo bem.

Mas agora os zumbi-istas postulam que essa coisa misteriosa não faz nada, então você precisa de uma nova explicação para o porquê de você dizer que é consciente.

Isso não é vitalismo. Isso é algo tão bizarro que os vitalistas cuspiriam seus cafés. "Quando os fogos queimam, eles liberam flogisto. Mas o flogisto não tem nenhum impacto experimentalmente detectável em nosso universo, então você terá que procurar uma explicação separada de por que um fogo pode derreter a neve." O quê?

Os dualistas de propriedade têm a impressão de que, se postularem uma nova força ativa, algo que tenha um impacto causal nos observáveis, estarão se arriscando demais?

Eu diria que se você postula uma propriedade misteriosa, separada, adicional e inerentemente mental da consciência, acima e além das posições e velocidades, então, nesse ponto, você já se arriscou o máximo que pode. Postular essa coisa da consciência e depois postular ainda que ela não faz nada - pelo amor dos gatinhos fofos, por quê?

Não há nem mesmo um motivo de carreira óbvio. "Oi, sou um filósofo da consciência. Meu assunto é a coisa mais importante do universo e eu deveria receber muito financiamento? Bem, é gentil da sua parte dizer isso, mas, na verdade, o fenômeno que estudo não faz absolutamente nada." (O argumento do impacto na carreira é inválido, mas eu o digo para deixar uma linha de retirada.)

Chalmers critica o dualismo de substância com base no fato de que é difícil ver qual nova teoria da física, qual nova substância que interage com a matéria, poderia possivelmente explicar a consciência. Mas o dualismo de propriedade tem o mesmo problema. Não importa sobre que tipo de propriedade dual você fale, como exatamente ela explica a consciência?

Quando Chalmers postulou uma propriedade extra que é a consciência, ele deu um salto sobre o inexplicável. Como isso ajuda sua teoria a especificar ainda que essa propriedade extra não tem efeito? Por que não apenas deixá-la ser causal?

Se eu fosse ser cruel, este seria o momento de trazer o dragão - mencionar a parábola de Carl Sagan sobre o dragão na garagem. "Eu tenho um dragão na minha garagem." Ótimo! Quero vê-lo, vamos lá! "Você não pode vê-lo - é um dragão invisível." Oh, eu gostaria de ouvi-lo então. "Desculpe, é um dragão inaudível." Eu gostaria de medir sua produção de dióxido de carbono. "Ele não respira." Vou jogar um saco de farinha no ar, para delinear sua forma. "O dragão é permeável à farinha."

Um motivo para tentar tornar sua teoria infalsificável é que, no fundo, você teme colocá-la à prova.

Sir Roger Penrose (físico) e Stuart Hameroff (neurologista) são dualistas de substância; eles pensam haver algo misterioso acontecendo no quântico, que Everett está errado e que o "colapso da função de onda" é fisicamente real, e que é aí que a consciência reside e como ela exerce efeito causal sobre seus lábios quando você diz em voz alta "Penso, logo existo". Acreditando nisso, eles previram que os neurônios se protegeriam da decoerência por tempo suficiente para manter estados quânticos macroscópicos.

Isso está em processo de teste, e até agora, as perspectivas não parecem boas para Penrose -

- mas a conduta básica de Penrose é cientificamente respeitável. Não Bayesiana, talvez, mas ainda fundamentalmente saudável. Ele apresentou uma hipótese maluca. Ele disse como testá-la. Ele saiu e tentou realmente testá-la.

Como eu disse uma vez a Stuart Hameroff: "Acho que a hipótese que você está testando é completamente desesperadora, e seus experimentos definitivamente devem ser financiados. Mesmo que você não encontre exatamente o que está procurando, você está procurando em um lugar onde ninguém mais está, e você pode encontrar algo interessante."

Então, uma rejeição desagradável do epifenomenalismo seria que os zumbi-istas têm medo de dizer que a coisa da consciência pode ter efeitos, porque então os cientistas poderiam procurar as propriedades extras e não as encontrar.

Eu não acho que isso seja realmente verdade para Chalmers, no entanto. Se Chalmers não tivesse auto-honestidade, ele poderia tornar as coisas muito mais fáceis para si mesmo.

(Mas caso Chalmers esteja lendo isso e tenha medo da falsificação, apontarei que se o epifenomenalismo for falso, então há alguma outra explicação para aquilo que chamamos de consciência, e ela será eventualmente encontrada, deixando a teoria de Chalmers em ruínas; então, se Chalmers se importa com seu lugar na história, ele não tem motivo para endossar o epifenomenalismo[30] a menos que ele realmente pense que é verdadeiro.)

Chalmers é um dos filósofos mais frustrantes que conheço. Às vezes me pergunto se ele está fazendo um Ateísmo Conquistado. Chalmers faz essa análise realmente clara… e então vira à esquerda no último minuto. Ele expõe tudo o que há de errado com o cenário do Mundo Zumbi e, após reduzir todo o argumento a pedaços, calmamente o aceita.

Chalmers faz a mesma coisa quando expõe, com calma e detalhes, o problema de dizer que nossas próprias crenças na consciência são justificadas, quando nossos gêmeos zumbis dizem a mesma coisa pelas mesmas razões e estão errados.

Na teoria de Chalmers, o fato de Chalmers dizer que acredita na consciência não pode ser justificado causalmente; a crença não é causada pelo próprio fato. Na ausência de consciência, Chalmers escreveria os mesmos artigos pelas mesmas razões.

No epifenomenalismo, o fato de Chalmers dizer que acredita na consciência não pode ser justificado como o produto de um processo que sistematicamente produz crenças verdadeiras, porque o gêmeo zumbi escreve os mesmos artigos usando o mesmo processo sistemático e está errado.

Chalmers admite isso. Chalmers, na verdade, explica o argumento em grande detalhe em seu livro. Ok, então Chalmers provou solidamente que ele não está justificado em acreditar na consciência epifenomenal, certo? Não. Chalmers escreve:

A experiência consciente está no centro do nosso universo epistêmico; temos acesso a ela diretamente. Isso levanta a questão: o que justifica nossas crenças sobre nossas experiências, se não é um vínculo causal com essas experiências, e se não são os mecanismos pelos quais as crenças são formadas? Acho que a

30 NT. **Epifenomenalismo** é a doutrina que sustenta serem os estados mentais meros subprodutos de processos físicos, sem exercer influência causal sobre o mundo material. Em outras palavras, a atividade cerebral geraria sensações e pensamentos, porém estes não poderiam modificar a própria atividade cerebral nem afetar eventos externos.

resposta para isso é clara: é ter as experiências que justifica as crenças. Por exemplo, o próprio fato de eu ter uma experiência vermelha agora fornece justificativa para minha crença de que estou tendo uma experiência vermelha...

Como meu gêmeo zumbi não tem experiências, ele está em uma situação epistêmica muito diferente da minha, e seus julgamentos não têm a justificativa correspondente. Pode ser tentador objetar que se minha crença reside no reino físico, sua justificativa deve residir no reino físico; mas isso é um *non sequitur*[31]. Do fato de que não há justificativa no reino físico, pode-se concluir que a porção física de mim (meu cérebro, digamos) não está justificada em sua crença. Mas a questão é se estou justificado na crença, não se meu cérebro está justificado na crença, e se o dualismo de propriedade estiver correto, então há mais em mim do que meu cérebro.

Então - se entendi bem essa tese - há um núcleo seu, acima e além do seu cérebro, que acredita que não é um zumbi e experimenta diretamente não ser um zumbi; e, portanto, suas crenças são justificadas.

Mas Chalmers acabou de escrever tudo isso em seu livro muito físico, e o Chalmers zumbi também.

O Chalmers zumbi não pode ter escrito o livro devido ao "eu" central do zumbi acima do cérebro; deve haver alguma razão inteiramente diferente, nas leis da física.

Segue-se que, mesmo que haja uma parte de Chalmers escondida que seja consciente e acredite na consciência, diretamente e sem mediação, existe também um subespaço separável de Chalmers, um subsistema cognitivo causalmente fechado que opera inteiramente dentro da física. Esse "eu exterior" é o responsável pela narrativa interna de Chalmers e pelos artigos que escreve sobre consciência.

Eu não vejo nenhuma maneira de evitar a acusação de que, na própria teoria de Chalmers, esse Chalmers externo separável é perturbado. Esta é a parte de Chalmers que é a mesma neste mundo ou no Mundo Zumbi; e em ambos os mundos escrevem artigos de filosofia sobre consciência sem nenhuma razão válida. Os artigos de filosofia de Chalmers não são produzidos por aquele núcleo interno de consciência e crença-na-consciência; eles são produzidos pela mera física da narrativa interna que faz os dedos de Chalmers tocarem as teclas de seu computador.

E ainda assim, esse Chalmers exterior perturbado está escrevendo artigos de filosofia que por acaso estão perfeitamente certos, por um milagre separado e adicional. Não um milagre logicamente necessário (então o Mundo Zumbi não seria logicamente possível). Um milagre fisicamente contingente, que acontece de ser verdade no que pensamos ser nosso universo, mesmo que a ciência nunca possa distinguir nosso universo do Mundo Zumbi.

Ou pelo menos, essa parece ser a implicação do que o Chalmers exterior auto-confessadamente perturbado está nos dizendo.

Acho que falo por todos os reducionistas quando digo "Hein?"

Isso não são epiciclos[32]. Isso é: "Os movimentos planetários seguem esses epiciclos - mas os epiciclos não fazem nada de fato - há algo mais que faz os planetas se moverem da mesma forma que os epiciclos dizem que deveriam, o que eu não consegui explicar - e a propósito, eu diria isso mesmo se não houvesse epiciclos."

Tenho uma perspectiva não convencional sobre filosofia porque vejo tudo com o objetivo de projetar uma IA; especificamente, uma Inteligência Artificial Geral auto-aprimorável com estrutura motivacional estável.

---

31 NT. ***Non sequitur*** (latim para "não se segue") é uma falácia lógica em que a conclusão não decorre validamente das premissas apresentadas. Em outras palavras, o desfecho do argumento carece de conexão lógica com o que foi afirmado antes.

32 NT. **Epiciclos**: Órbitas circulares menores utilizadas na astronomia ptolomaica para descrever o movimento planetário em torno da Terra. Esses círculos adicionais buscavam explicar variações na trajetória aparente dos planetas, como o movimento retrógrado. O modelo reflete o esforço de harmonizar observações celestes com o paradigma geocêntrico predominante na Antiguidade.

Quando penso em projetar uma IA, pondero princípios como a teoria da probabilidade, a noção Bayesiana de evidência como diagnóstico diferencial e, acima de tudo, a coerência reflexiva. Qualquer IA auto-modificável que comece em um estado reflexivamente inconsistente não permanecerá assim por muito tempo.

Se uma inteligência artificial (IA) que pode se modificar analisa uma parte dela mesma que diz "B" sempre que a condição A é verdadeira, e a IA percebe que essa parte dela tende a escrever informações erradas na memória, ela identificou o que parece um bug. Então, a IA se consertará para não escrever "B" quando a condição A for verdadeira.

Qualquer teoria epistemológica que desconsidere a coerência reflexiva não é uma boa teoria para usar na construção de IA auto-aprimorável. Este é um argumento decisivo da minha perspectiva, considerando o que pretendo realmente usar a filosofia. Então tenho que inventar uma teoria reflexivamente coerente de qualquer maneira. E quando faço isso, caramba, a coerência reflexiva acaba fazendo sentido intuitivo.

Então essa é a maneira incomum com que eu tendo a pensar sobre essas coisas. E agora olho de volta para Chalmers:

O "Chalmers exterior" causalmente fechado (que não é influenciado de forma alguma pelo "Chalmers interior" que tem consciência e crenças adicionais separadas) deve estar realizando alguma operação sistematicamente não confiável e injustificada que de alguma forma inexplicável faz com que a narrativa interna produza crenças sobre um "Chalmers interior" que estão corretas sem nenhuma razão lógica no que acontece de ser o nosso universo.

Mas não há garantia possível para o Chalmers exterior ou qualquer IA auto-inspetora reflexivamente coerente acreditar nessa correção misteriosa. Um bom design de IA deveria, penso eu, parecer uma inteligência reflexivamente coerente incorporada em um sistema causal, com uma teoria testável de como esse mesmo sistema causal produz crenças sistematicamente precisas no caminho para alcançar seus objetivos.

Então, a IA examinará Chalmers e verá um sistema cognitivo causal fechado produzindo uma narrativa interna que está proferindo bobagens. Bobagens que parecem ter um alto impacto no que Chalmers pensa que deveria ser considerado uma pessoa moralmente valiosa.

Este não é um problema necessário para os teóricos da IA Amigável. É um problema apenas se você for um epifenomenalista. Se você acredita nos reducionistas (a consciência acontece nos átomos) ou nos dualistas de substância (a consciência é uma coisa imaterial causalmente potente), as pessoas que falam sobre consciência estão falando sobre algo real, e uma IA Bayesiana reflexivamente consistente pode ver isso rastreando a cadeia de causalidade do que faz as pessoas dizerem "consciência".

De acordo com Chalmers, o sistema cognitivo causalmente fechado da narrativa interna de Chalmers está (misteriosamente) funcionando mal de uma forma que, não por necessidade, mas apenas em nosso universo, acontece milagrosamente de estar correta. Além disso, a narrativa interna afirma que "a narrativa interna está misteriosamente funcionando mal, mas acontece milagrosamente de estar ecoando corretamente os pensamentos justificados do núcleo interno epifenomenal", e novamente, em nosso universo, acontece milagrosamente de estar correta.

Ah, qual é!

Não deveria haver um ponto em que você simplesmente desiste de uma ideia? Onde, em algum nível intuitivo cru, você simplesmente pensa: O que diabos eu estava pensando?

A humanidade acumulou uma ampla experiência com a aparência de teorias corretas sobre o mundo. Esta não é a aparência de uma teoria correta.

"Argumento da incredulidade", você diz. Tudo bem, você quer que seja explicado? A referida teoria Chalmeriana postula múltiplos milagres complexos inexplicáveis. Isso reduz sua probabilidade prévia, pela regra da conjunção da probabilidade e a Navalha de Occam. Portanto, é dominada por pelo menos duas teorias que postulam menos milagres, a saber:

- **Dualismo de substância:**
- Existe uma coisa da consciência que ainda não é compreendida, uma coisa super física extraordinária que afeta visivelmente nosso mundo; e essa coisa é o que nos faz falar sobre consciência.

- **Reducionismo não baseado na fé:**
- Aquilo que chamamos de "consciência" acontece na física, de uma forma ainda não compreendida, assim como aconteceu nas últimas três mil vezes em que a humanidade se deparou com algo misterioso.
- Sua intuição de que nenhuma substância material pode possivelmente somar à consciência está incorreta. Se você realmente soubesse exatamente por que fala sobre consciência, isso lhe daria novos insights, de uma forma que você não pode antecipar agora; e depois você perceberia que seus argumentos sobre a física normal não ter espaço para a consciência eram falhos.

Compare com:

- **Dualismo de propriedade epifenomenal:**
- A matéria tem propriedades adicionais de consciência que ainda não são compreendidas. Essas propriedades são epifenomenais em relação à física ordinariamente observável - elas não fazem diferença no movimento das partículas.
- Separadamente, existe uma razão ainda não compreendida na física normal pela qual os filósofos falam sobre consciência e inventam teorias de propriedades duais.
- Milagrosamente, quando os filósofos falam sobre consciência, as leis de ligação do nosso mundo estão exatamente corretas para tornar essa conversa sobre consciência correta, mesmo que ela surja de um mau funcionamento (elaboração de conclusões logicamente injustificadas) no sistema cognitivo causalmente fechado que digita artigos de filosofia.

Sei que estou falando de uma experiência limitada aqui. Mas com base na minha experiência limitada, o Argumento Zumbi pode ser um candidato à ideia mais perturbada de toda a filosofia.

Há momentos em que, como racionalista, você tem que acreditar em coisas que parecem estranhas para você. A relatividade parece estranha, a mecânica quântica parece estranha, a seleção natural parece estranha.

Mas essas estranhezas são sustentadas por evidências maciças. Há uma diferença entre acreditar em algo estranho porque a ciência o confirmou esmagadoramente -

- versus acreditar em uma proposição que parece completamente perturbada, devido a um argumento filosófico grande e complicado centrado em milagres não especificados e pontos cegos gigantes que nem mesmo se afirma serem compreendidos -
- em um caso em que, mesmo que você aceite tudo o que lhe foi dito até agora, depois o fenômeno ainda parecerá um mistério e ainda terá a mesma qualidade de impenetrabilidade maravilhosa que tinha no início.

A coisa correta para um racionalista dizer neste ponto, se todos os argumentos de David Chalmers parecem individualmente plausíveis, o que não me parece, é:

> Ok... eu não sei como a consciência funciona... admito isso... e talvez eu esteja abordando todo o problema de forma errada, ou fazendo as perguntas erradas... mas essa coisa de zumbi não pode estar certa. Os argumentos não são suficientemente sólidos para me fazer acreditar nisso - especialmente quando o aceitar não me fará sentir menos confuso. Em um nível central, isso simplesmente não parece ser a maneira como a realidade poderia realmente funcionar[33].

33 NT. Texto original em inglês. *Okay... I don't know how consciousness works... I admit that ...and maybe I'm approaching the whole problem wrong, or asking the wrong questions... but this zombie business can't possibly be right. The arguments aren't nailed down enough to make me believe this—especially when accepting it won't make me feel any less confused. On a core gut level, this just doesn't look like the way reality could really really work.*

Note que não estou dizendo que isso é um substituto para uma refutação analítica cuidadosa da tese de Chalmers. O Sistema 1 não é um substituto para o Sistema 2, embora ajude a apontar o caminho. Você ainda precisa rastrear onde estão os problemas especificamente.

Chalmers escreveu um grande livro, nem todo disponível através da prévia gratuita do Google. Eu não dupliquei as longas cadeias de argumentos onde Chalmers expõe os argumentos contra si em detalhes calmos. Eu apenas tentei adicionar uma refutação final à última defesa apresentada por Chalmers, que Chalmers ainda não respondeu ao meu conhecimento. Devolvi a bola para a quadra dele, por assim dizer.

Mas, sim, em um nível central, a coisa sensata a fazer quando você vê a conclusão do argumento zumbi é dizer “Isso não pode estar certo” e começar a procurar uma falha.

Referências

[1] Chalmers, The Conscious Mind.

222 - Respostas zumbis

Estou um pouco cansado hoje, pois fiquei acordado até as 3 da manhã escrevendo o ensaio de mais de 6.000 palavras sobre zumbis de ontem. Então, hoje apenas responderei ao Richard e resolverei uma ponta solta que notei no dia seguinte.

(A) Richard Chappell escreve:

> Uma nota terminológica (para evitar confusões desnecessárias): o que você chama de "concebível", outros de nós chamaríamos apenas de "aparentemente concebível[34].

A lacuna entre "ainda não vejo uma contradição" e "isso é logicamente possível" é tão grande (é NP-completo até em alguns casos aparentemente simples) que você realmente deveria ter duas palavras diferentes. Como o argumento dos zumbis é impulsionado enquanto essa enorme lacuna pode ser varrida para debaixo do tapete de pequenas diferenças terminológicas, eu realmente acho que seria uma boa ideia dizer "concebível" versus "logicamente possível" ou talvez até ter uma distinção ainda mais visível. Não posso escolher terminologia profissional que já foi estabelecida, mas, em um caso como este, eu poderia seriamente me recusar a usá-la.

Talvez eu diga "aparentemente concebível" para o tipo de informação que os defensores dos zumbis obtêm imaginando Mundos Zumbis, e "logicamente possível" para o tipo de informação estabelecida exibindo um modelo completo ou prova lógica. Observe o tamanho da lacuna entre a informação que você pode obter fechando os olhos e imaginando zumbis, e a informação que você precisa para sustentar o argumento do epifenomenalismo.

> Ou seja, sua visão seria caracterizada como uma forma de materialismo do Tipo-A, a visão de que zumbis não são nem mesmo (genuinamente) concebíveis, muito menos metafisicamente possíveis[35].

O materialismo Tipo-A é um grande conjunto; você não deve atribuir esse conjunto a mim até me ver concordar com cada uma das partes. Acho que alguém que pergunta "O que é consciência?" está fazendo uma pergunta legítima, tem uma demanda legítima por um insight; não necessariamente penso que a resposta tome a forma de "Aqui está esta coisa que tem todas as propriedades que você atribuiria à consciência, por tal e tal razão", mas pode, em certa medida, consistir em insights que fazem você perceber que estava fazendo a pergunta da maneira errada.

Isso não é ser eliminativo sobre a consciência. É ser realista sobre o tipo de insights a esperar, diante de um problema que (1) parece que deve ter alguma solução, (2) parece que não pode ter nenhuma solução, e (3) está sendo discutido de uma maneira que depende muito da arquitetura ad-hoc da cognição humana, que não é totalmente compreendida.

(1) Até onde posso dizer, você não identificou nenhuma contradição lógica na descrição do mundo zumbi. Você apenas apontou ser meio estranho. Mas existem muitos mundos possíveis bizarros por aí. Isso

34 NT. Texto original em inglês. *A terminological note (to avoid unnecessary confusion): what you call "conceivable," others of us would merely call "apparently conceivable."*

35 NT. Texto original em inglês. *That is, your view would be characterized as a form of Type-A materialism, the view that zombies are not even (genuinely) conceivable, let alone metaphysically possible.*

não é motivo para supor uma contradição implícita. Então ainda é completamente misterioso para mim qual é essa suposta contradição.

Ok, explicarei do ponto de vista materialista:

1. O mundo zumbi, por definição, contém todas as partes do nosso mundo que estão no fecho da relação "causado por" ou "efeito de" qualquer fenômeno observável. Em particular, contém a causa de eu visivelmente dizer: "Penso, logo existo."
2. Quando foco minha consciência interior na minha consciência interior, logo em seguida experimento minha narrativa interna dizendo: "Estou focando minha consciência interior na minha consciência interior," e posso, se escolher, dizer isso em voz alta.
3. Intuitivamente, parece que minha consciência interior está causando minha narrativa interna a dizer certas coisas, e que minha narrativa interna pode fazer meus lábios dizerem certas coisas.
4. A palavra "consciência," se tiver algum significado, refere-se àquilo-que-é ou àquilo-que-causa ou àquilo-que-me-faz-dizer-que-tenho-consciência-interna.
5. De (3) e (4) seguiria que, se o mundo zumbi está fechado em relação às causas de eu dizer "Penso, logo existo," o mundo zumbi contém aquilo a que nos referimos como "consciência."
6. Por definição, o mundo zumbi não contém consciência.
7. (3) parece-me ter uma probabilidade bastante alta de ser empiricamente verdadeiro. Portanto, avalio uma alta probabilidade empírica de que o mundo zumbi é logicamente impossível.

Você pode salvar o Mundo Zumbi deixando a causa da minha narrativa interna dizendo "Penso, logo existo" ser algo completamente diferente da consciência. Em conjunto com a suposição de que a consciência existe, esta é a parte que me pareceu desorientada.

Mas se o que foi dito acima é concebível, então o Mundo Zumbi não é concebível?

Não, porque as duas construções do Mundo Zumbi envolvem dar à palavra "consciência" diferentes referentes empíricos, como "água" em nosso mundo significando H_2O versus "água" na Terra Gêmea de Putnam significando XYZ. Para o Mundo Zumbi ser logicamente possível, não basta que, pelo que você sabia sobre como o mundo empírico funcionava, a palavra "consciência" poderia ter se referido a um epifenômeno que é totalmente diferente da consciência que conhecemos. O Mundo Zumbi carece de consciência, não de "consciência"—é um mundo sem H_2O, não um mundo sem "água." Isso é o que é necessário para sustentar a afirmação empírica, "Você poderia eliminar o referente do que é significado por 'consciência' do nosso mundo, mantendo todos os átomos no mesmo lugar."

Ou seja: Sustento que é um fato empírico, dado o que a palavra "consciência" realmente se refere, sendo logicamente impossível eliminar a consciência sem mover nenhum átomo. O que significaria eliminar "consciência" de um mundo, em vez de consciência, eu não especularei.

(2) É enganoso dizer que é "milagroso" (na visão do dualista de propriedades) que nossos *qualia* se alinhem tão perfeitamente com o mundo físico. Há uma lei natural que garante isso, afinal. Então, não é mais milagroso do que qualquer outra necessidade nômica logicamente contingente (por exemplo, as constantes em nossas leis físicas).

É a própria lei natural que é "milagrosa"—conta como um elemento adicional complexo-improvável da teoria a ser postulada, sem ter sido ela própria justificada em termos de coisas já conhecidas. Postula-se (a) um mundo interior que é consciente, (b) um mundo exterior disfuncional que fala sobre consciência sem razão, e (c) que os dois se alinham perfeitamente. A afirmação (c) não decorre de (a) e (b), e, portanto, é um postulado separado.

Concordo que esse uso de "milagroso" conflita com o sentido filosófico de violar uma lei natural; quis dizer no sentido de improbabilidade aparecendo de nenhuma fonte aparente, à la crença em movimento perpétuo. Portanto, a palavra foi mal escolhida no contexto. Mas não é intuitivamente o tipo de coisa que deveríamos chamar de milagre? Sua consciência não faz realmente você dizer que está consciente, há uma coisa física separada que faz você dizer que está consciente, mas também há uma lei que alinha os dois—isso é realmente um evento em uma ordem semelhante de estranheza a uma hóstia assumindo a substância da

carne de Cristo enquanto possui a aparência exata e o comportamento externo de uma hóstia, há apenas uma lei natural que garante isso, sabe.

> Ou seja, o Chalmers Zumbi (ou "Exterior") não conclui nada, porque suas falas são sem sentido. Com razão ainda mais forte, ele não conclui nada indevidamente. Ele está apenas emitindo sons; estes não são mais suscetíveis de avaliação epistêmica do que os piados de um pássaro.

Analisando isso do ponto de vista do design da IA, é possível construir uma IA que melhore continuamente um componente interno de si mesma que se correlacione (no sentido de informações mútuas ou relações sistemáticas) com o ambiente, talvez incluindo números de ponto flutuante de um tipo que eu chamaria de "probabilidades", pois seguiriam as relações internas exigidas pelos Teoremas de Cox quando a IA encontrasse novas informações, ou seja, novos estímulos sensoriais.

Você dirá que, a menos que a IA seja mais do que meros transistores - a menos que tenha o duplo aspecto - a IA não tem crenças.

Acho que minhas opiniões sobre isso foram expressas de maneira bastante clara em "A Verdade Simples".

Para mim, parece bastante direto construir mapas que se correlacionem com territórios de maneiras sistemáticas, sem mencionar nada além de coisas de pura causalidade física. A IA gera um mapa do Texas. Outra IA voa com o mapa até o Texas e verifica se as rodovias estão nos lugares correspondentes, emitindo "Verdadeiro" quando detecta uma correspondência e "Falso" quando detecta uma discrepância. Você pode se recusar a chamar isso de "um mapa do Texas", mas as próprias IAs continuam emitindo "Verdadeiro" ou "Falso", e as referidas IAs vão emitir "Falso" quando olharem para a crença de Chalmers em um núcleo epifenomênico, e eu, por exemplo, concordaria com elas.

Está claro que a função de mapear a realidade é realizada estritamente pelo Chalmers Exterior. Todo o processo de produzir representações de crença é tratado pela estrutura bayesiana nas interações causais. Não sobra nada para o Chalmers Interior fazer, exceto abençoar todo o assunto com um significado epifenomênico. Agora, "significado" é algo inteiramente não relacionado à correspondência sistemática mapa-território ou à capacidade de usar esse mapa para navegar na realidade. Então, quando se trata de falar sobre "precisão", muito menos "precisão sistemática", parece-me que deveríamos conseguir determinar isso estritamente olhando para o Chalmers Exterior.

(B) No texto de ontem, deixei de fora uma suposição quando escrevi:

> Quando uma IA auto-modificável analisa uma parte de si mesma que conclui "B" sob a condição A (ou seja, uma parte que escreve "B" na memória sempre que a condição A é verdadeira), e a IA inspeciona essa parte, determina como ela opera causalmente no contexto do universo maior e decide que essa parte tende sistematicamente a escrever dados falsos na memória, então a IA encontrou o que parece um bug, e a IA se auto-modificará para não escrever "B" no pool de crenças sob a condição A.
>
> …

Mas não há nenhuma justificativa possível para que o Chalmers exterior ou qualquer IA auto-inspecionadora reflexivamente coerente acredite nessa correção misteriosa. Um bom projeto de IA deve, eu acho, ser uma inteligência reflexivamente coerente com uma teoria testável de como opera como um sistema causal, portanto, com uma teoria testável de como esse sistema causal produz crenças sistematicamente precisas no caminho para alcançar seus objetivos.

Na verdade, você precisa de uma suposição adicional para o que foi dito acima, que é que um "bom projeto de IA" (o tipo que eu estava pensando, de qualquer forma) julga sua própria racionalidade de maneira modular; ela impõe a racionalidade global ao impor a racionalidade local. Se há uma peça que, em relação ao seu contexto, é localmente sistematicamente não confiável—para algumas possíveis crenças "B_i" e condições "A_i", adiciona alguns "B_i" ao pool de crenças sob a condição local "A_i", onde a reflexão pelo sistema indica que B_i é falso (ou, no caso de crenças probabilísticas, não é preciso) quando a condição local "A_i" é verdadeira—então isso é um bug. Esse tipo de modularidade é uma maneira de tornar o problema tratável, e é como

penso atualmente sobre o design de IA de primeira geração. [Edição de 2013: A noção real que eu tinha em mente aqui já foi desenvolvida e formalizada em *Tiling Agents for Self-Modifying AI*, (Agentes Modulares para IA Auto-Adaptativa) seção 6].

A noção é que um sistema cognitivo causalmente fechado—como uma IA projetada por seus programadores para usar apenas partes causalmente eficazes; ou uma IA cuja teoria de seu próprio funcionamento é inteiramente testável; ou o Chalmers exterior que escreve artigos de filosofia—que acredita que tem um eu interior epifenomênico, deve estar fazendo algo sistematicamente não confiável porque concluiria a mesma coisa em um Mundo Zumbi. Uma mente cujas partes são sistematicamente localmente confiáveis, em relação aos seus contextos, seria sistematicamente globalmente confiável. Portanto, uma mente que é globalmente não confiável deve conter pelo menos uma parte localmente não confiável. Então, um sistema cognitivo causalmente fechado inspecionando a si mesmo em busca de confiabilidade local deve descobrir que pelo menos um passo envolvido na adição da crença em um eu interior epifenomênico é não confiável.

Se houver outras maneiras de mentes serem reflexivamente coerentes que evitem essa prova de descrença e zumbis, os filósofos estão convidados a tentar especificá-las.

A razão pela qual tenho que especificar tudo isso é que, de outra forma, você obtém um tipo de coerência reflexiva extremamente barata onde a IA nunca pode se rotular como não confiável. Por exemplo, se a IA encontrar uma parte de si mesma que calcula 2 + 2 = 5 (no contexto circundante de contar ovelhas), a IA raciocinará: "Bem, essa parte funciona mal e diz que 2 + 2 = 5… mas por pura coincidência, 2 + 2 é igual a 5, ou assim parece para mim… então, embora a parte pareça sistematicamente não confiável, é melhor mantê-la como está, ou ela lidará com esse caso especial errado." É por isso que falo sobre impor confiabilidade global impondo confiabilidade sistemática local - se você apenas comparar suas crenças globais com suas crenças globais, você não vai a lugar nenhum.

Isso tem uma lição geral: Mostre que seus argumentos são globalmente confiáveis por virtude de cada passo ser localmente confiável; não apenas compare as conclusões dos argumentos com suas intuições. [Edição de 2013: Consulte *Proofs, Implications, and Models* (Provas, Impicações e Modelos) para uma discussão sobre o fato de que a lógica válida é localmente válida].

(C) Um seguidor anônimo escreveu:

> Um ponto secundário, este, mas acredito que sua etimologia para "n'shama" está errada. Está relacionada à palavra para "respiração", não "ouvir". A raiz de "ouvir" contém um ayin, que n'shama não contém[36].

Agora, isso é o que chamo de uma coincidência milagrosamente enganosa - embora a palavra *N'Shama*[37] tenha surgido por razões completamente diferentes, soava exatamente da maneira certa para me fazer pensar que se referia a um ouvinte interior.

Opa!.

---

36 NT. Texto original em inglês. *A sidepoint, this, but I believe your etymology for "n'shama" is wrong. It is related to the word for "breath," not "hear." The root for "hear" contains an ayin, which n'shama does not.*

37 NT. ***N'Shama*** (em hebraico, נשמה) refere-se, de modo geral, à "alma" ou "essência vital". Em textos religiosos e místicos, pode também ser traduzido como "sopro" de vida, enfatizando o aspecto espiritual e transcendente desse termo na tradição judaica.

223 - O princípio generalizado anti-zumbi

Cada problema que resolvi tornou-se uma regra que serviu posteriormente para resolver outros problemas[38].

— René Descartes, Discurso do Método [1]

"Zumbis" são seres hipotéticos que são idênticos a nós humanos, átomo por átomo, governados por todas as mesmas leis físicas visíveis a terceiros, exceto que eles não são conscientes.

Embora a filosofia seja complicada, o argumento central contra zumbis é simples: quando você foca sua consciência interior em sua própria consciência, sua narrativa interna (aquela vozinha na sua cabeça que fala seus pensamentos) diz "Estou ciente de estar ciente" logo depois. Então você diz isso em voz alta e depois digita em um teclado, criando uma postagem visível para terceiros.

A consciência, seja ela uma substância, um processo, ou um nome para uma confusão, não é epifenomênica; sua mente pode capturar o ouvinte interior em ação e dizer isso em voz alta. O fato de eu ter digitado este parágrafo parece pelo menos refutar a ideia de que a consciência não tem consequências experimentalmente detectáveis.

Odeio dizer "Então, vamos aceitar isso e seguir em frente" sobre uma questão tão controversa filosoficamente, mas parece que uma considerável maioria dos comentadores do Overcoming Bias aceitam isso. E há outras conclusões que você só pode alcançar após aceitar que você não pode subtrair a consciência e deixar o universo parecendo o mesmo. Então proponho aceitarmos isso e seguir em frente.

A forma do Argumento Anti-Zumbi parece que deveria se generalizar, tornando-se um Princípio Anti-Zumbi. Mas qual é a generalização adequada?

Digamos, por exemplo, que alguém diz: "Tenho um interruptor na minha mão, que não afeta seu cérebro de forma alguma; e se esse interruptor for acionado, você deixará de estar consciente." O Princípio Anti-Zumbi descarta isso também, com a mesma estrutura de argumento?

Parece-me que, no caso acima, a resposta é sim. Em particular, você pode dizer: "Mesmo depois de seu interruptor ser acionado, eu ainda falarei sobre consciência pelos mesmos motivos que antes. Se estou consciente agora, ainda estarei consciente após você acionar o interruptor."

Os filósofos podem objetar: "Mas agora você está equiparando a consciência com falar sobre consciência! E o Mestre Zumbi, o chatbot que regurgita um corpus remixado de discurso humano amador sobre consciência?"

Mas eu não equiparei "consciência" com comportamento verbal. A premissa central é que, entre outras coisas, o verdadeiro referente de "consciência" é também a causa de os humanos falarem sobre ouvintes internos.

38 NT. Texto original em inglês. *Each problem that I solved became a rule which served afterwards to solve other problems.*

Como argumentei (extensamente) na sequência sobre palavras, o que você deseja ao definir uma palavra nem sempre é uma definição aristotélica perfeita, necessária e suficiente; às vezes você só quer um mapa do tesouro que leve você ao referente extensional. Portanto, "aquilo que de fato me faz falar sobre uma consciência indizível" não é uma definição necessária e suficiente. Mas se o que de fato me faz discursar sobre uma consciência indizível não é "consciência", então…

… Então o discurso fica bastante fútil. Isso não é um argumento decisivo contra zumbis—uma questão empírica não pode ser resolvida por meras dificuldades de discurso. Mas se você tentar desafiar o Princípio Anti-Zumbi, terá problemas com o significado do seu discurso, não apenas com sua plausibilidade.

Poderíamos definir a palavra "consciência" para significar "o que quer que de fato faça os humanos falarem sobre 'consciência'"? Isso teria a vantagem poderosa de garantir que há pelo menos um fato nomeado pela palavra "consciência". Mesmo se nossa crença na consciência for uma confusão, "consciência" nomearia a arquitetura cognitiva que gerou a confusão. Mas estabelecer uma definição é apenas prometer usar uma palavra consistentemente; não resolve nenhuma questão empírica, se nossa consciência interior nos faz falar sobre nossa consciência interior.

Proponho voltarmos ao Interruptor de Desligar.

Se permitirmos que o Argumento Anti-Zumbi se aplique contra o Interruptor de Desligar, então o Princípio Generalizado Anti-Zumbi não diz apenas: "Qualquer mudança que não seja detectável experimentalmente (IPED)[39] não pode remover sua consciência." O acionamento do interruptor é detectável experimentalmente, mas ainda parece altamente improvável que remova sua consciência.

Talvez o Princípio Anti-Zumbi diga: "Qualquer mudança que não te afete de maneira IPED não pode remover sua consciência"?

Mas é uma estipulação razoável dizer que acionar o interruptor não te afeta de maneira IPED? Todas as partículas no interruptor estão interagindo com as partículas que compõem seu corpo e cérebro. Há efeitos gravitacionais—pequenos, mas reais e IPED. A atração gravitacional de um interruptor de um grama a dez metros de distância é de cerca de m/s^2. Isso é cerca de metade do diâmetro de um nêutron por segundo a cada segundo, bem abaixo do ruído térmico, mas muito acima do nível de Planck.

Poderíamos acionar o interruptor a anos-luz de distância, caso em que o acionamento não teria efeito causal imediato sobre você (seja lá o que "imediato" significa neste caso) (se o Modelo Padrão da física estiver correto).

Mas não parece que deveríamos alterar o experimento mental dessa maneira. Parece que, se um interruptor desconectado for acionado do outro lado da sala, você não deveria esperar que seu ouvinte interior apagasse como uma lâmpada, porque o interruptor "obviamente não muda" aquilo que é a verdadeira causa de você falar sobre um ouvinte interior. Seja o que for que você realmente seja, você não espera que o interruptor interfira nisso.

Isso é um grande passo.

Se você nega que é um passo razoável, é melhor nunca chegar perto de um interruptor novamente. Mas ainda assim, é um grande passo.

A ideia central do reducionismo é que nossos mapas do universo são multiníveis para economizar poder de computação, mas a física parece ser estritamente de um único nível. Todo nosso discurso sobre o universo ocorre usando referências muito acima do nível das partículas fundamentais.

O acionamento do interruptor muda as partículas fundamentais do seu corpo e cérebro. Ele as empurra por diâmetros inteiros de nêutrons para longe de onde estariam de outra forma.

Na vida cotidiana, nós disfarçamos uma mudança tão pequena dizendo que o interruptor "não te

39 NT. Da sigla em inglês para ***in-principle experimentally detectable.***

afeta". Mas ele te afeta. Muda tudo por diâmetros inteiros de nêutrons! O que poderia estar permanecendo o mesmo? Apenas a descrição que você daria dos níveis superiores de organização—células, proteínas, impulsos viajando ao longo de um axônio neural. Como o mapa é muito menos detalhado que o território, ele deve mapear muitos estados diferentes para a mesma descrição.

Qualquer tipo razoável de descrição humana do cérebro que fale sobre neurônios e padrões de atividade (ou até mesmo as conformações dos microtúbulos individuais que compõem axônios e dendritos) não mudará quando você acionar um interruptor do outro lado da sala. Núcleos são maiores que nêutrons, átomos são maiores que núcleos, e quando você chega ao nível molecular, aquela pequena força gravitacional desapareceu da lista de coisas que você se preocupa em rastrear.

Mas se você somar muito pequenas forças gravitacionais, elas eventualmente te puxarão pela sala e te despedaçarão por forças de maré, então claramente um pequeno efeito não é "nenhum efeito".

Talvez a força de maré de uma pequena tração possa, por uma coincidência incrível, puxar um único íon de cálcio ainda mais perto de um canal iônico. Isso faz com que o íon seja puxado um pouco mais cedo, levando um único neurônio a disparar infinitesimalmente mais cedo do que faria normalmente. Essa diferença amplifica-se caoticamente, resultando em um impulso neural inteiro que não ocorreria de outra forma. Isso pode levar a um trem de pensamento diferente, que pode desencadear um ataque epiléptico, que o mata, fazendo com que você deixe de estar consciente…

Se você somar muitos efeitos quantitativos pequenos, obtém um grande efeito quantitativo—grande o suficiente para bagunçar qualquer coisa que você queira nomear. E assim, afirmar que o interruptor tem literalmente zero efeito nas coisas que você se importa, é exagerar.

Mas com apenas um interruptor, a força exercida é muito menor que as incertezas térmicas, sem falar das incertezas quânticas. Se você não espera que sua consciência pisque dentro e fora de existência como resultado da oscilação térmica, então certamente não deveria esperar apagar como uma lâmpada quando alguém espirra a um quilômetro de distância.

O alerta bayesiano notará que acabei de fazer um argumento sobre expectativas, estados de conhecimento, crenças justificadas sobre o que pode e não pode desligar sua consciência.

Isso não destrói necessariamente o Argumento Anti-Zumbi. Probabilidades não são certezas, mas as leis da probabilidade são teoremas; se a racionalidade diz que você não pode acreditar em algo com suas informações atuais, então isso é uma lei, não uma sugestão.

Ainda assim, esta versão do Argumento Anti-Zumbi é mais fraca. Não tem o status claro e absolutamente inequívoco de: "Você não pode eliminar a consciência enquanto deixa todos os átomos exatamente no mesmo lugar." (Ou substitua "todos os átomos" por "todas as causas com efeitos experimentalmente detectáveis em princípio" e "mesma função de onda" por "mesmo lugar", etc.)

Mas a nova versão do Argumento Anti-Zumbi ainda se mantém. Você pode dizer: "Não sei o que realmente é a consciência e suspeito que possa estar fundamentalmente confuso sobre a questão. Mas se a palavra se refere a algo, refere-se a algo que é, entre outras coisas, a causa de eu falar sobre a consciência. Agora, eu não sei por que falo sobre a consciência. Mas isso acontece dentro do meu crânio e eu espero que tenha algo a ver com neurônios disparando. Ou talvez, se eu entendesse realmente a consciência, teria que falar sobre um nível ainda mais fundamental, como microtúbulos ou neurotransmissores difundindo-se por meio de um canal sináptico. Mas ainda assim, esse interruptor que você acabou de acionar tem um efeito nos meus neurotransmissores e microtúbulos que é muito, muito menor do que o ruído térmico a 310 Kelvin. Então, seja qual for a verdadeira causa de eu falar sobre a consciência, não espero que seja muito afetada pela atração gravitacional desse interruptor. Talvez seja um pouquinho afetada? Mas certamente não apagará como uma lâmpada. Espero continuar falando sobre a consciência quase exatamente da mesma maneira depois disso, quase exatamente pelos mesmos motivos."

Esta aplicação do Princípio Anti-Zumbi é mais fraca. Mas é também muito mais geral. E, em termos de puro bom senso, correta.

O reducionista e o dualista de substância têm, na verdade, duas versões diferentes da declaração acima. O reducionista ainda diz: "O que quer que me faça falar sobre a consciência, parece provável que as partes importantes ocorram em um nível funcional muito mais alto do que os núcleos atômicos. Alguém que entendesse a consciência poderia abstrair dos neurônios individuais disparando e falar sobre arquiteturas cognitivas de alto nível, e ainda descrever como minha mente produz pensamentos como 'Penso, logo existo.' Então, mover as coisas pelo diâmetro de um núcleon não deveria afetar minha consciência (exceto talvez com uma probabilidade muito pequena, ou por uma quantidade muito pequena, ou não até depois de um atraso significativo)."

O dualista de substância ainda diz: "O que quer que me faça falar sobre a consciência, tem que ser algo além da física computacional que conhecemos, o que significa que pode muito bem envolver efeitos quânticos. Mas ainda assim, minha consciência não pisca dentro e fora sempre que alguém espirra a um quilômetro de distância. Se isso acontecesse, eu notaria. Seria como pular alguns segundos, ou sair de uma anestesia geral, ou às vezes dizer, 'Não penso, logo não sou.' Então, como é um fato físico que as vibrações térmicas não perturbam a essência da minha consciência, não espero que acionar o interruptor a perturbe também."

De qualquer forma, você não deve esperar que seu senso de consciência desapareça quando alguém diz a palavra "Abracadabra," mesmo que isso tenha algum efeito físico infinitesimal no seu cérebro -

Mas espere! Se você ouvir alguém dizer a palavra "Abracadabra," isso tem um efeito muito perceptível no seu cérebro - tão grande que até o seu cérebro pode notá-lo. Isso pode alterar sua narrativa interna; você pode pensar: "Por que essa pessoa acabou de dizer 'Abracadabra'?"

Bem, mas ainda assim você espera continuar falando sobre a consciência quase exatamente da mesma maneira, quase exatamente pelos mesmos motivos.

E, novamente, não é que "consciência" esteja sendo equiparada a "aquilo que faz você falar sobre a consciência." É apenas que a consciência, entre outras coisas, faz você falar sobre a consciência. Então, qualquer coisa que faça sua consciência apagar como uma lâmpada deve fazer você parar de falar sobre a consciência.

Se fizermos algo com você, onde você não vê como isso poderia possivelmente mudar sua narrativa interna - aquela vozinha na sua cabeça que às vezes diz coisas como "Penso, logo existo," cujas palavras você pode escolher dizer em voz alta - então isso não deveria fazer você deixar de estar consciente.

E isso é verdade mesmo se a narrativa interna for apenas "quase a mesma" e as causas dela também forem praticamente as mesmas; entre as causas que são praticamente as mesmas está o que quer que você signifique por "consciência."

Se você está se perguntando para onde tudo isso está indo e por que é importante ponderar por tanto tempo um Princípio Generalizado Anti-Zumbi aparentemente óbvio, então considere o seguinte debate:

Albert: "Suponha que eu substitua todos os neurônios da sua cabeça por pequenos neurônios artificiais robóticos que tenham as mesmas conexões, o mesmo comportamento de entrada e saída local e regras internas de estado e aprendizado análogas."

Bernice: "Isso é me matar! Não haveria mais um ser consciente ali."

Charles: "Bem, ainda haveria um ser consciente ali, mas não seria eu."

Sir Roger Penrose: "O experimento mental que você propõe é impossível. Você não pode duplicar o comportamento dos neurônios sem explorar a gravidade quântica. Dito isso, não há muito sentido em eu continuar participando desta conversa." (Afastando-se.)

Albert: "Suponha que a substituição seja realizada um neurônio de cada vez, e a troca ocorra tão rapidamente que não faça diferença para o processamento global."

Bernice: "Como isso é possível?"

Albert: "O pequeno robô nada até o neurônio, o envolve, escaneia, aprende a duplicá-lo e, em seguida, assume o comportamento de repente, entre um impulso e o próximo. Na verdade, a imitação é tão boa que seu comportamento externo é o mesmo que seria se o cérebro fosse deixado intocado. Talvez não o mesmo, mas o impacto causal é muito menor do que o ruído térmico a 310 Kelvin."

Charles: "E daí?"

Albert: "Então suas crenças não violam o Princípio Generalizado Anti-Zumbi? O que quer que tenha acontecido, não mudou sua narrativa interna! Você continuará falando sobre a consciência exatamente pelos mesmos motivos de antes."

Bernice: "Esses pequenos robôs são um Mestre Zumbi. Eles me farão falar sobre a consciência, mesmo que eu não esteja consciente. O Mundo Zumbi é possível se você permitir que haja um Mestre Zumbi adicional, extra, experimentalmente detectável—o que esses robôs são."

Charles: "Oh, isso não está certo, Bernice. Os pequenos robôs não estão tramando como falsificar a consciência, ou processando um corpus de texto de amadores humanos. Eles estão fazendo a mesma coisa que os neurônios fazem, só que em silício em vez de carbono."

Albert: "Espere, você não acabou de concordar comigo?"

Charles: "Eu nunca disse que a nova pessoa não seria consciente. Eu disse que não seria eu."

Albert: "Bem, obviamente o Princípio Anti-Zumbi é generalizado para dizer que essa operação não perturbou a verdadeira causa de você falar sobre essa coisa de 'eu'."

Charles: "Uh-uh! Sua operação certamente perturbou a verdadeira causa de eu falar sobre a consciência. Substituiu uma causa diferente em seu lugar, os robôs. Agora, só porque essa nova causa também acontece de ser consciente—fala sobre a consciência pela mesma razão generalizada—não significa que seja a mesma causa que estava lá originalmente."

Albert: "Mas eu nem precisaria te contar sobre a operação dos robôs. Você não notaria. Se você pensa, com base em evidências introspectivas, que você é, em um sentido importante, 'a mesma pessoa' que era cinco minutos atrás, e eu faço algo com você que não muda as evidências introspectivas disponíveis, então sua conclusão de que você é a mesma pessoa que era cinco minutos atrás deve ser igualmente justificada. O Princípio Generalizado Anti-Zumbi não diz que se faço algo com você que altera sua consciência, muito menos faz de você uma pessoa completamente diferente, você deveria notar de alguma forma?"

Bernice: "Não se você me substituir por um Mestre Zumbi. Então não há ninguém lá para notar."

Charles: "A introspecção não é perfeita. Muitas coisas acontecem dentro do meu cérebro que eu não noto."

Albert: "Você está postulando fatos epifenomênicos sobre consciência e identidade!"

Bernice: "Não estou! Posso detectar experimentalmente a diferença entre neurônios e robôs."

Charles: "Não estou! Posso detectar experimentalmente o momento em que o velho eu é substituído por uma nova pessoa."

Albert: "Sim, e eu posso detectar o acionamento do interruptor! Você está detectando algo que não faz uma diferença perceptível para a verdadeira causa de seu discurso sobre consciência e identidade pessoal. E a prova é que você continuará falando da mesma forma depois."

Bernice: "Isso é devido ao seu Mestre Zumbi robótico!"

Charles: "Só porque duas pessoas falam sobre 'identidade pessoal' por motivos semelhantes não as

torna a mesma pessoa."

Acho que o Princípio Generalizado Anti-Zumbi apoia a posição de Albert, mas as razões terão que esperar por ensaios futuros. Preciso de outros pré-requisitos e, além disso, este ensaio já está muito longo.

Mas você vê a importância da pergunta: "Até que ponto você pode generalizar o Argumento Anti--Zumbi e ele ainda ser válido?"

A composição das futuras civilizações galácticas pode depender da resposta.

Referências

[1] René Descartes, Discours de la Méthode, vol. 45 (Librairie des Bibliophiles, 1887).

224 - GAZP x GLUT

Em *The Unimagined Preposterousness of Zombies* (A Absurda Improbabilidade dos Zumbis), Daniel Dennett diz: [1]

> Até o momento, vários filósofos me disseram que planejam aceitar meu desafio de oferecer uma defesa não questionadora de zumbis, mas o único que vi até agora envolve postular um ser "logicamente possível", mas fantástico - um descendente da fantasia da *Giant Lookup Table* de Ned Block…[40]

Uma "*Giant Lookup Table*" (Tabela de Consulta Gigante), em linguagem de programador, é quando você implementa uma função como uma tabela gigante de entradas e saídas, geralmente para economizar no tempo de computação. Se meu programa precisa saber o produto multiplicativo de duas entradas entre 1 e 100, posso escrever um algoritmo de multiplicação que calcula cada vez que a função é chamada, ou posso pré-calcular uma Tabela de Consulta Gigante com 10.000 entradas e dois índices. Há momentos em que você quer fazer isso, embora não para multiplicação - momentos em que você vai reutilizar a função muitas vezes e ela não tem muitas entradas possíveis; ou quando os ciclos de clock são baratos durante a inicialização, mas muito caros durante a execução.

Tabelas de Consulta Gigantes ficam muito grandes, muito rapidamente. Uma GLUT de todas as possíveis conversas de vinte réplicas com dez palavras por observação, usando apenas 850 palavras do inglês básico, exigiria $7{,}6 \times 10^{585}$ entradas.

Substituir um cérebro humano por uma Tabela de Consulta Gigante de todas as entradas sensoriais e saídas motoras possíveis (em relação a algum esquema de digitização refinado) exigiria uma quantidade irracionalmente grande de armazenamento de memória. Mas "em princípio", como os filósofos gostam de dizer, isso poderia ser feito.

A GLUT não é um zumbi no sentido clássico, porque é micro-fisicamente diferente de um humano. (Na verdade, uma GLUT não pode realmente funcionar na mesma física que um humano; é muito grande para caber em nosso universo. Para fins filosóficos, proponho ignorarmos isso e supor um suprimento ilimitado de armazenamento de memória.)

Mas a GLUT é um zumbi? Ou seja, ela se comporta exatamente como um humano sem estar consciente?

A língua do corpo da GLUT fala sobre consciência. Seus dedos escrevem artigos de filosofia. De todas as maneiras, desde que você não olhe no crânio, a GLUT parece exatamente como um humano... o que certamente parece um exemplo válido de um zumbi: se comporta como um humano, mas não há ninguém em casa.

A menos que a GLUT seja consciente, caso em que não seria um exemplo válido.

Não me lembro de ter visto alguém afirmar que uma GLUT é consciente. (Admito que minha leitura

---

40 NT. Texto original em inglês. *To date, several philosophers have told me that they plan to accept my challenge to offer a non-question-begging defense of zombies, but the only one I have seen so far involves postulating a "logically possible" but fantastic being—a descendent of Ned Block's Giant Lookup Table fantasy…*

nesta área não está no nível profissional; sinta-se à vontade para me corrigir.) Mesmo pessoas acusadas de serem (eca!) funcionalistas não afirmam que as GLUTs podem ser conscientes.

As GLUTs são o *reducio ad absurdum* para qualquer um que sugira que a consciência é simplesmente um padrão de entrada-saída, eliminando assim todas as preocupações problemáticas sobre o que acontece por dentro.

Então, o que o Princípio Anti-Zumbi Generalizado (GAZP) diz sobre a Tabela de Consulta Gigante (GLUT)?

À primeira vista, parece que uma GLUT é o próprio arquétipo de um Mestre Zumbi - um sistema distinto, adicional, detectável e não consciente que anima um zumbi e o faz falar sobre consciência por razões diferentes.

No interior da GLUT, há apenas um software muito simples que procura entradas e recupera saídas. Até mesmo falar sobre um "software simples" é exagerar, em um caso como este. Uma GLUT é mais como ROM do que uma CPU. Poderíamos igualmente falar sobre uma série de trilhos comutados pelos quais algumas bolas rolam para fora de uma pilha previamente armazenada e para uma calha - ponto final; isso é tudo o que a GLUT faz.

Um porta-voz do Povo pelo Tratamento Ético dos Zumbis responde: "Ah, é isso que todos os anti-mecanicistas dizem, não é? Que quando você olha no cérebro, você só encontra um monte de neurotransmissores abrindo canais iônicos? Se os canais iônicos podem ser conscientes, por que não as alavancas e as bolas rolando em caixas?"

"O problema não é as alavancas", responde o funcionalista, "o problema é que uma GLUT tem o padrão errado de alavancas. Você precisa de alavancas que implementem coisas como, digamos, formação de crenças sobre crenças, ou auto-modelagem... Poxa, você precisa da capacidade de escrever coisas na memória apenas para o tempo poder passar para o cálculo. A menos que você pense que é possível programar um ser consciente em Haskell[41]*."*

"Eu não sei sobre isso", diz o porta-voz do PETZ, "tudo o que sei é que este suposto zumbi escreve artigos filosóficos sobre consciência. De onde vêm esses artigos de filosofia, se não da consciência?"

Boa pergunta! Ponderemos sobre isso profundamente.

Há um jogo na física chamado Siga a Energia. O pai de Richard Feynman jogou com o jovem Richard:

> Era o tipo de coisa que meu pai teria falado: "O que faz isso funcionar? Tudo funciona porque o Sol está brilhando." E então nos divertiríamos discutindo isso:
>
> "Não, o brinquedo funciona porque a mola está enrolada", eu diria.
>
> "Como a mola foi enrolada?", ele perguntaria.
>
> "Eu a enrolei."
>
> "E como você se moveu?"
>
> "Comendo."
>
> "E a comida cresce apenas porque o Sol está brilhando. Então é porque o Sol está brilhando que todas essas coisas estão se movendo." Isso transmitiria o conceito de que o movimento é simplesmente a transformação

41 NT. **Haskell** é uma linguagem de programação puramente funcional e estaticamente tipada, conhecida por sua ênfase em **imutabilidade**, **avaliação preguiçosa** (lazy evaluation) e **sistema de tipos avançado**. Criada nos anos 1990, é amplamente usada em ambientes acadêmicos e industriais para explorar conceitos como alto nível de abstração, segurança de tipos e programação concorrente.

do poder do Sol[42]. [2]

Quando você fica um pouco mais velho, aprende que a energia é conservada, nunca criada ou destruída, então a noção de usar energia não faz muito sentido. Você nunca pode mudar a quantidade total de energia, então em que sentido você a está usando?

Então, quando os físicos crescem, eles aprendem a jogar um novo jogo chamado Siga a Negentropia - o qual é realmente o mesmo jogo que eles estavam jogando o tempo todo; apenas as regras são mais matemáticas, o jogo é mais útil e os princípios são mais difíceis de entender conceitualmente.

Os racionalistas aprendem um jogo chamado Siga a Improbabilidade, a versão adulta de "Como Você Sabe?" A regra do jogo do racionalista é que toda crença que parece improvável precisa de uma quantidade equivalente de evidências para justificá-la. (Este jogo tem regras incrivelmente semelhantes à Siga a Negentropia.)

Sempre que alguém viola as regras do jogo do racionalista, você pode encontrar um lugar em seu argumento onde uma quantidade de improbabilidade aparece do nada; e isso é tanto um sinal de problema quanto, digamos, um projeto engenhoso de rodas e engrenagens ligadas que se mantém funcionando para sempre.

Alguém vem até você e diz: "Eu acredito com fé firme e duradoura que há um objeto no cinturão de asteroides, de trinta centímetros de diâmetro e composto inteiramente de bolo de chocolate; você não pode provar que isso é impossível". Mas, a menos que a pessoa tivesse acesso a algum tipo de evidência para essa crença, seria altamente improvável que uma crença correta se formasse espontaneamente. Então ou a pessoa pode apontar evidências, ou a crença não será verdadeira. "Mas você não pode provar que é impossível para minha mente gerar espontaneamente uma crença que por acaso está correta!" Não, mas esse tipo de geração espontânea é altamente improvável, assim como, digamos, um ovo que se quebra sozinho.

Em Siga a Improbabilidade, é altamente suspeito até mesmo falar sobre uma hipótese específica sem ter evidências suficientes para reduzir o espaço de hipóteses possíveis. Por que você não está dando atenção igual a um decilhão de outras hipóteses igualmente plausíveis? Você precisa de evidências suficientes para encontrar a hipótese do "bolo de chocolate no cinturão de asteroides" no espaço de hipóteses - caso contrário, não há razão para dar a ela mais atenção do que a um trilhão de outros candidatos como "Há uma cômoda de madeira no cinturão de asteroides" ou "O Monstro do Espaguete Voador vomitou nos meus tênis".

Em "Siga a Improbabilidade", você não tem permissão para tirar grandes hipóteses específicas complicadas do nada sem já ter uma quantidade correspondente de evidências; porque não é realista supor que você poderia espontaneamente começar a discutir a hipótese verdadeira por pura coincidência.

Um filósofo diz: "O crânio deste zumbi contém uma Tabela de Consulta Gigante de todas as entradas e saídas para o cérebro de algum humano". Esta é uma improbabilidade muito grande. Então você pergunta: "Como esse evento improvável ocorreu? De onde veio a GLUT?"

Agora, este não é um procedimento filosófico padrão para experimentos mentais. No procedimento filosófico padrão, você tem permissão para postular coisas como "Suponha que você estivesse cavalgando um raio de luz..." sem se preocupar com a possibilidade física, muito menos com a mera improbabilidade. Mas neste caso, a origem da GLUT importa; e é por isso que é importante entender a pergunta motivadora: "De onde veio a improbabilidade?"

A resposta óbvia é que você pegou uma especificação computacional de um cérebro humano e a usou para pré-calcular a Tabela de Consulta Gigante. (Criando assim incontáveis googols de seres humanos,

42 NT. Texto original em inglês. *It was the kind of thing my father would have talked about: "What makes it go? Everything goes because the Sun is shining." And then we would have fun discussing it: "No, the toy goes because the spring is wound up," I would say. "How did the spring get wound up?" he would ask. "I wound it up." "And how did you get moving?" "From eating." "And food grows only because the Sun is shining. So it's because the Sun is shining that all these things are moving." That would get the concept across that motion is simply the transformation of the Sun's power.*

alguns deles em extrema dor, a super maioria enlouquecida em um universo de caos onde as entradas não têm relação com as saídas. Mas dane-se a ética, isso é para a filosofia.)

Nesse caso, a GLUT está escrevendo artigos sobre consciência devido a um algoritmo consciente. A GLUT não é um zumbi, assim como um celular não é um zumbi porque pode falar sobre consciência enquanto é apenas um pequeno dispositivo eletrônico de consumo. O celular está apenas transmitindo discursos de filosofia de quem quer que esteja do outro lado da linha. Uma GLUT gerada a partir de uma especificação de cérebro originalmente humano está fazendo a mesma coisa.

"Tudo bem", diz o filósofo, "a GLUT foi gerada aleatoriamente e por acaso tem as mesmas relações de entrada-saída que algum humano de referência."

Como, exatamente, você gerou aleatoriamente a GLUT?

"Usamos uma fonte de aleatoriedade verdadeira - um dispositivo quântico."

Mas um dispositivo quântico apenas implementa a instrução Ramificar em Ambas as Direções; quando você gera um bit de uma fonte de aleatoriedade quântica, o resultado determinístico é que um conjunto de ramos do universo (nuvens de amplitude localmente conectadas) vê 1, e outro conjunto de universos vê 0. Faça isso 4 vezes, crie 16 (conjuntos de) universos.

Então, na verdade, isso é como dizer que você obteve a GLUT escrevendo todas as sequências possíveis de 0s e 1s do tamanho de uma GLUT, em uma caixa realmente enorme de tabelas de consulta; e então colocando a mão na caixa e, de alguma forma, retirando uma GLUT que por acaso correspondia à especificação de um cérebro humano. De onde veio a improbabilidade?

Porque se isso não foi apenas uma coincidência, se você teve alguma função de acesso a caixa que retirasse especificamente uma GLUT correspondente a um humano intencionalmente, e não por acaso, então essa função provavelmente seria consciente. Nesse caso, a GLUT seria novamente um celular, não um zumbi. Estaria conectada a um humano a duas remoções, em vez de uma, mas ainda seria um ser celular. Essa seria uma tentativa engenhosa de esconder a fonte da improbabilidade.

Agora veja onde Siga a Improbabilidade nos levou: onde está a fonte da língua deste corpo falando sobre um ouvinte interior? A consciência não está na tabela de consulta. A consciência não está na fábrica que fabrica muitas tabelas de consulta possíveis. A consciência estava no que apontou para uma tabela de consulta específica já fabricada e disse: "Use essa!"

Você pode ver por que introduzi o jogo Siga a Improbabilidade. Normalmente, quando estamos conversando com uma pessoa, tendemos a pensar que o que está no crânio deve ser "onde está a consciência". É apenas jogando Siga a Improbabilidade que podemos perceber que a verdadeira fonte da conversa que estamos tendo é aquilo responsável pela improbabilidade da conversa - por mais distante no tempo ou no espaço, como o Sol movendo um brinquedo de corda.

"Não, não!" diz o filósofo. "No experimento mental, eles não estão gerando aleatoriamente muitas GLUTs e depois usando um algoritmo consciente para escolher uma GLUT que parece humana! Estou especificando que, neste experimento mental, eles colocam a mão na caixa de GLUTs inconcebivelmente vasta e, por puro acaso, retiram uma GLUT que é idêntica às entradas e saídas de um cérebro humano! Pronto! Eu te encurralei agora! Você não pode mais jogar Siga a Improbabilidade!"

Oh! Então sua especificação é a fonte da improbabilidade aqui.

Quando jogamos Siga a Improbabilidade novamente, acabamos fora do experimento mental, olhando para o filósofo.

Aquilo que aponta para a única GLUT que fala sobre consciência, em todo o vasto espaço de possibilidades, é agora… a pessoa consciente nos pedindo para imaginar todo esse cenário. E nossos próprios cérebros, que preencherão o espaço em branco quando imaginarmos: "O que essa GLUT dirá em resposta a 'Fale sobre seu ouvinte interior'?"

A moral desta história é que quando você segue o discurso sobre "consciência", geralmente encontra consciência. Nem sempre está bem na sua frente. Às vezes está muito bem escondida. Mas está lá. Daí o Princípio Anti-Zumbi Generalizado.

Se houver um Mestre Zumbi na forma de um chatbot que processa e remixa o discurso humano amador sobre "consciência", os humanos que geraram o corpus de texto original são conscientes.

Se algum dia você entender a consciência e olhar para trás e ver que há um programa que você pode escrever que produzirá um discurso filosófico confuso que soa muito parecido com humanos sem ser consciente - então, quando eu perguntar "Como este software chegou a soar semelhante aos humanos?" a resposta é que você o escreveu para soar semelhante aos humanos conscientes, em vez de escolher com base no critério de semelhança com outra coisa. Isso não significa que seu pequeno Mestre Zumbi seja consciente - mas significa que posso encontrar consciência em algum lugar do universo rastreando a cadeia de causalidade, o que significa que não estamos inteiramente no Mundo Zumbi.

Mas suponha que alguém tenha realmente colocado a mão em uma caixa de GLUTs e, por puro acaso, retirado uma GLUT que escreveu artigos de filosofia?

Bem, então ela não seria consciente. Na minha humilde opinião.

Quero dizer, tem que haver mais do que entradas e saídas.

Caso contrário, até mesmo uma GLUT seria consciente, certo?

Ah, e para aqueles de vocês que estão se perguntando como esse tipo de coisa se relaciona com meu trabalho diário…

Nesse ramo de atividade, você encontra um número enorme de pessoas que acreditam que uma IA poderosa, gerada arbitrariamente, será "moral". Elas não conseguem concordar entre si sobre o porquê ou o que querem dizer com a palavra "moral"; mas todas concordam que a teoria da IA Amigável é desnecessária. Ao perguntar a essas pessoas como uma IA gerada arbitrariamente produz resultados morais, elas oferecem racionalizações elaboradas com base no que consideram "moral"; e existem vários tipos de problemas com essa abordagem, mas o principal deles é, "Você tem certeza de que a IA seguiria a mesma linha de pensamento que você inventou para argumentar sobre a moral humana, quando, ao contrário de você, a IA não começa sabendo o que você quer que ela racionalize?" Você poderia chamar o contra-princípio de "Siga a Informação da Decisão" ou algo semelhante. Você pode explicar uma IA que faz coisas improvavelmente boas dizendo-me como escolheu o projeto da IA a partir de um enorme espaço de possibilidades, mas, de outra forma, a improbabilidade está sendo tirada do nada - embora cada vez mais disfarçada, pois as premissas racionalizadas são racionalizadas por sua vez.

Então, eu já fiz uma série completa de ensaios que eu mesmo gerei usando Siga a Improbabilidade. Mas eu não expliquei as regras explicitamente naquela época, porque eu ainda não havia feito os ensaios sobre termodinâmica…

Só pensei em mencionar isso. É incrível como muitos dos meus ensaios coincidentemente acabam incluindo ideias surpreendentemente relevantes para a discussão da teoria da IA Amigável… se você acredita em coincidência.

Referências

[1] Daniel C. Dennett, "The Unimagined Preposterousness of Zombies," Journal of Consciousness Studies 2 (4 1995): 322–26.

[2] Richard P. Feynman, "Judging Books by Their Covers," in Surely, You're Joking, Mr. Feynman! (New York: W. W. Norton & Company, 1985).

225 - Crença no invisível implícito

Uma lição generalizada que não se deve tirar do Argumento Anti-Zumbi é: "Tudo que você não pode ver não existe."

É tentador concluir essa regra geral. Tornaria o Argumento Anti-Zumbi muito mais simples, em ocasiões futuras, se pudéssemos tomar isso como premissa. Mas, infelizmente, isso simplesmente não é bayesiano.

Suponha que eu transmita um fóton em direção ao infinito, não mirando em nenhuma estrela ou galáxia, apontando-o para um dos grandes vazios entre superaglomerados. Baseando-me na física padrão, em outras palavras, não espero que este fóton intercepte algo em seu caminho. O fóton está se movendo na velocidade da luz, então não posso persegui-lo e capturá-lo novamente.

Se a expansão do universo está acelerando, como sustenta a cosmologia atual, chegará um ponto futuro em que não espero conseguir interagir com o fóton nem mesmo em princípio — um tempo futuro além do qual não espero que o cone de luz futuro do fóton intercepte minha linha do mundo. Mesmo que uma espécie alienígena capturasse o fóton e corresse de volta para nos contar, eles não poderiam viajar rápido o suficiente para compensar a expansão acelerada do universo.

Devo acreditar que, no momento em que não posso mais interagir com ele, nem mesmo em princípio, o fóton desaparece?

Não.

Isso violaria a Conservação de Energia. E a Segunda Lei da Termodinâmica. E praticamente todas as outras leis da física. E provavelmente as Três Leis da Robótica. Implicaria que o fóton sabe que me importo com ele e sabe exatamente quando desaparecer.

É uma ideia tola.

Mas se você pode acreditar na existência contínua de fótons que se tornaram experimentalmente indetectáveis para você, por que isso não implica uma licença geral para acreditar no invisível?

(Se você quiser pensar sobre esta questão por conta própria, faça-o antes de continuar lendo…)

Embora eu não tenha conseguido encontrar uma fonte no Google, lembro-me de ter lido que quando foi proposto pela primeira vez que a Via Láctea era nossa galáxia - que o rio nebuloso de luz no céu noturno era composto por milhões (ou até bilhões) de estrelas - a Navalha de Occam foi invocada contra a nova hipótese. Porque, veja bem, a hipótese multiplicava vastamente o número de "entidades" no universo acreditado. Ou talvez tenha sido a sugestão de que as "nebulosas" — aquelas manchas nebulosas vistas por meio de um telescópio — poderiam ser galáxias cheias de estrelas, que provocou a invocação da Navalha de Ocam.

Lex parsimoniae: Entia non sunt multiplicanda praeter necessitatem[43]

43 NT. Latim. Lei da Parcimônia: As entidades não devem ser multiplicadas além da necessidade. A frase, associada ao filósofo medieval Guilherme de Ockham ("Navalha de Ocam"), defende a priorização de explicações ou modelos mais simples que evitem premissas supérfluas.

Essa foi a formulação original de Ocam, a lei da parcimônia: as entidades não devem ser multiplicadas além da necessidade.

Se você postula bilhões de estrelas em que ninguém acreditou antes, você está multiplicando entidades, não está?

Não. Existem duas formalizações bayesianas da Navalha de Ocam: a indução de Solomonoff[44] e o Comprimento Mínimo de Mensagem. Nenhuma penaliza as galáxias por serem grandes.

E é melhor que não o façam! Uma das lições da história é que aquilo que chamamos de realidade continua se revelando cada vez maior e mais vasta. Lembra quando a Terra estava no centro do universo? Lembra quando ninguém havia inventado o número de Avogadro[45]? Se a Navalha de Ocam estivesse pesando contra a multiplicação de entidades todas as vezes, teríamos que começar a duvidar da Navalha de Occam, porque ela teria consistentemente se mostrado errada.

Na indução de Solomonoff, a complexidade do seu modelo é a quantidade de código no software que você tem que escrever para simular seu modelo. A quantidade de código, não a quantidade de RAM que ele usa ou o número de ciclos que leva para computar. Um modelo do universo que contém bilhões de galáxias contendo bilhões de estrelas, cada estrela feita de um bilhão de trilhões de decilhões de quarks, vai precisar de muita RAM para rodar — mas o código só precisa descrever o comportamento dos quarks, e as estrelas e galáxias podem ser deixadas para funcionar por conta própria. Estou falando de forma semi-metafórica aqui - há coisas no universo além de quarks - mas o ponto é que postular um bilhão extra de galáxias não conta contra o tamanho do seu código, se você já descreveu uma galáxia. Só precisa de um pouco mais de RAM, e a Navalha de Ocam não se importa com RAM.

Por quê não? O formalismo do Comprimento Mínimo de Mensagem, o qual é quase equivalente à indução de Solomonoff, pode tornar o princípio mais claro: se você tem que contar a alguém como seu modelo do universo funciona, você não precisa especificar individualmente a localização de cada quark em cada estrela em cada galáxia. Você só precisa escrever algumas equações. A quantidade de "coisas" que obedece à equação não afeta quanto tempo leva para escrever a equação. Se você codificar a equação em um arquivo, e o arquivo tiver 100 bits de comprimento, então há 2^{100} outros modelos que teriam aproximadamente o mesmo tamanho de arquivo, e você precisará de cerca de 100 bits de evidência de suporte. Você tem uma quantidade limitada de massa de probabilidade; e a priori, você tem que dividir essa massa entre todas as mensagens que poderia enviar; e assim, postular um modelo em um espaço de modelo de 2^{100} alternativas significa que você tem que aceitar uma penalidade de probabilidade prévia de 2^{-100} - mas ter mais galáxias não acrescenta a isso.

Postular bilhões de estrelas em bilhões de galáxias não afeta o comprimento da mensagem que descreve o comportamento geral de todas essas galáxias. Assim, você não sofre uma penalidade de probabilidade por ter as mesmas equações descrevendo mais coisas. (Desde que os sucessos preditivos do seu modelo não sejam sensíveis às condições iniciais exatas. Se você tiver que especificar as posições exatas de todos os quarks para que seu modelo preveja tão bem quanto faz, os quarks extras contam como uma penalidade.)

Se você supõe que o fóton desaparece quando você não está mais olhando para ele, esta é uma lei adicional em seu modelo do universo. São as leis que são "entidades", custosas sob as leis da parcimônia. Quarks extras são gratuitos.

Então, isso se resume a: "Acredito que o fóton continua existindo enquanto voa para o nada, por que meus prioris dizem ser mais simples que ele continue existindo do que desapareça"?

---

44 NT. A **Indução de Solomonoff** é uma teoria fundamental na ciência da computação e na inteligência artificial que busca formalizar matematicamente o processo de inferência indutiva, combinando princípios da teoria da informação algorítmica, probabilidade e a Navalha de Ocam. Desenvolvida por Ray Solomonoff na década de 1960, ela propõe um método universal para prever sequências de dados futuros com base em observações passadas, assumindo que o ambiente segue uma distribuição probabilística computável, mas desconhecida.

45 NT. **Número de Avogadro** (símbolo: N_A) é uma constante fundamental da química e física que define a quantidade de entidades elementares (átomos, moléculas, íons etc.) contidas em **1 mol** de uma substância..

Foi o que pensei inicialmente, mas após reflexão, não é bem isso. (E não apenas porque isso abre a porta para abusos óbvios.)

Eu resumiria isso a uma distinção entre crença no invisível implícito e crença no invisível adicional.

Quando você acredita que o fóton continua existindo enquanto voa para o infinito, você não está acreditando nisso como um fato adicional.

O que você acredita (atribui probabilidade) é um conjunto de equações simples. Você acredita que essas equações descrevem o universo. Você acredita nessas equações porque são as mais simples que você pôde encontrar para descrever a evidência. Essas equações são altamente testáveis experimentalmente. Elas explicam enormes montanhas de evidências visíveis no passado e preveem os resultados de muitas observações no futuro.

Você acredita nessas equações, e é uma implicação lógica delas que o fóton continua existindo enquanto voa para o nada. Então você acredita nisso também.

Seus prioris, ou mesmo suas probabilidades, não falam diretamente sobre o fóton. O que você atribui probabilidade não é o fóton, mas as leis gerais. Quando você atribui probabilidade às leis da física como as conhecemos, você contribui automaticamente com essa mesma probabilidade para o fóton continuar existindo em seu caminho para o nada — se você acredita nas implicações lógicas daquilo em que acredita.

Não é que você acredite no invisível como tal, a partir do raciocínio sobre coisas invisíveis. Em vez disso, a evidência experimental suporta certas leis, e a crença nessas leis implica logicamente a existência de certas entidades com as quais você não pode interagir. Isso é crença no invisível implícito.

Por outro lado, se você acredita que o fóton é devorado pelo Monstro do Espaguete Voador e deixa de existir - talvez apenas nesta ocasião específica - ou mesmo se você acreditasse sem motivo que o fóton atingiu um grão de poeira em seu caminho - então você estaria acreditando em um evento invisível específico adicional, por si só. Se você pensasse que esse tipo de coisa acontece, em geral, você acreditaria em uma lei invisível específica adicional. Isso é crença no invisível adicional.

Para deixar claro por que às vezes você gostaria de pensar sobre invisíveis implícitos, suponha que você lançará uma nave espacial, quase na velocidade da luz, em direção a um superaglomerado distante. Quando a nave chegar lá e estabelecer uma colônia, a expansão do universo terá acelerado demais para que eles possam enviar uma mensagem de volta. Você considera que vale a pena o esforço puramente altruísta de estabelecer essa colônia, pelo bem de todas as pessoas que viverão lá e serão felizes? Ou você acha que a nave espacial desaparece antes de chegar lá? Isso poderia ser uma questão muito real em algum momento.

O assunto todo seria muito mais simples, admitidamente, se pudéssemos simplesmente descartar a existência de entidades com as quais não podemos interagir, de uma vez por todas — fazer o universo parar de existir na borda de nossos telescópios. Mas isso nos obrigaria a ser muito tolos.

Dizer que você nunca deveria precisar de uma crença separada e adicional sobre coisas invisíveis — que você só acredita em invisíveis que são implicações lógicas de leis gerais testáveis, e mesmo assim, não tem nenhuma crença adicional sobre eles que não seja implicação lógica de regras gerais visivelmente testáveis — parece realmente excluir todos os abusos de crença no invisível, quando aplicado corretamente.

Talvez eu devesse dizer: “você deve atribuir probabilidade prévia inalterada a invisíveis adicionais”, em vez de dizer “não acredite neles”. Mas se você pensa em uma crença como algo evidencialmente adicional, algo que você se preocupa em rastrear, algo onde você se preocupa em contar o apoio a favor ou contra, então é questionável se devemos alguma vez ter crenças adicionais sobre invisíveis adicionais.

Existem casos exóticos que quebram isso em teoria. (Por exemplo: os demônios epifenomenais estão observando você e torturarão $3 \uparrow\uparrow\uparrow 3$ vítimas por um ano, em algum lugar que você nunca poderá verificar o evento, se você alguma vez disser a palavra “Niblick”.) Mas não consigo pensar em um caso em que o princípio falhe na prática humana.

226 - Zumbis: O Filme

Fade in em torno de um grupo sério de oficiais militares uniformizados. Na cabeceira da mesa, um homem corpulento, General Fred, fala.

General Fred: Os relatórios estão confirmados. Nova York foi tomada… por zumbis.

Coronel Todd: De novo? Mas acabamos de ter uma invasão de zumbis 28 dias atrás!

General Fred: Esses zumbis… são diferentes. Eles são… zumbis filosóficos.

Capitão Mudd: Eles estão cheios de raiva, fazendo com que mordam as pessoas?

Coronel Todd: Eles perdem toda a capacidade de raciocínio?

General Fred: Não. Eles se comportam… exatamente como nós… exceto que não são conscientes.

(Silêncio domina a mesa.)

Coronel Todd: Meu Deus.

General Fred se move para um display computadorizado.

General Fred: Esta é Nova York, duas semanas atrás.

O display mostra multidões se movimentando pelas ruas, pessoas comendo em restaurantes, um caminhão recolhendo o lixo.

General Fred: Esta… é Nova York… agora.

O display muda, mostrando um trem de metrô lotado, um grupo de estudantes rindo em um parque e um casal de mãos dadas ao sol.

Coronel Todd: É pior do que eu imaginava.

Capitão Mudd: Como você sabe, exatamente?

Coronel Todd: Eu nunca vi nada tão brutalmente comum.

Um Cientista de jaleco de laboratório se levanta na extremidade da mesa.

Cientista: A doença zumbi elimina a consciência sem mudar o cérebro de forma alguma. Estamos tentando entender como a doença é transmitida. Nossa conclusão é que, como a doença ataca propriedades duais da matéria comum, ela deve, por si mesma, operar fora do nosso universo. Estamos lidando com um vírus epifenomênico.

General Fred: Você tem certeza?

Cientista: Tão certo quanto podemos estar na total ausência de evidências.

General Fred: Tudo bem. Compile um relatório sobre todos os epifenômenos já observados. O quê, onde e quem. Quero uma lista de tudo o que não aconteceu nos últimos cinquenta anos.

Capitão Mudd: Se o vírus é epifenomênico, como sabemos que ele existe?

Cientista: Da mesma forma que sabemos que somos conscientes.

Capitão Mudd: Ah, entendi.

General Fred: Os médicos fizeram algum progresso em encontrar uma cura epifenomênica?

Cientista: Eles tentaram todos os placebos do livro. Nada feito. Tudo o que fazem tem um efeito.

General Fred: Você trouxe um homeopata?

Cientista: Tentei, senhor! Não consegui encontrar nenhum!

General Fred: Excelente. E os taoistas?

Cientista: Eles se recusam a fazer qualquer coisa!

General Fred: Então ainda podemos ser salvos.

Coronel Todd: E David Chalmers? Ele não deveria estar aqui?

General Fred: Chalmers… foi uma das primeiras vítimas.

Coronel Todd: Oh, não.

(Corte para o interior de uma cela, completamente cercada por vidro reforçado, onde David Chalmers anda de um lado para o outro.)

Médico: David! David Chalmers! Você pode me ouvir?

Chalmers: Sim.

Enfermeira: Não adianta, doutor.

Chalmers: Estou perfeitamente bem. Refleti sobre minha consciência e não consigo detectar nenhuma diferença. Sei que seria esperado que eu dissesse isso, mas -

O médico se afasta da tela de vidro horrorizado.

Médico: Suas palavras, elas… não significam nada.

Chalmers: Isso é uma distorção grotesca das minhas visões filosóficas. Esse tipo de coisa não pode realmente acontecer!

Médico: Por que não?

Enfermeira: Sim, por que não?

Chalmers: Porque -

(Corte para dois Policiais, guardando uma estrada de terra que leva ao portão de aço imponente de um complexo de concreto gigantesco. Em seus uniformes, um distintivo lê Agência de Aplicação da Lei de Ligação.)

Policial 1: Você tem que tomar cuidado com esses bastardos inteligentes. Eles parecem humanos.

Eles falam como humanos. São idênticos aos humanos no nível atômico. Mas não são humanos.

Policial 2: Canalhas.

O enorme barulho de um motor pulsante ecoa sobre as colinas. Sobe o Homem em uma motocicleta branca. O Homem está usando óculos escuros pretos e um terno de couro preto com gravata de couro preta e botas de metal prateado. Sua barba branca flui ao vento. Ele para em frente ao portão.

Os Policiais se aproximam da motocicleta.

Policial 1: Declare seu negócio aqui.

Homem: É aqui que estão mantendo David Chalmers?

Policial 2: O que isso importa para você? Você é amigo dele?

Homem: Não posso dizer que sou. Mas até mesmo zumbis têm direitos.

Policial 1: Tudo bem, amigo, vamos ver suas qualia.

Homem: Eu não tenho nenhuma.

O Policial 2 de repente puxa uma arma, mantendo-a apontada para o Homem.

Policial 2: Aha! Um zumbi!

Policial 1: Não, zumbis afirmam ter qualia.

Policial 2: Então ele é um humano comum?

Policial 1: Não, eles também afirmam ter qualia.

Os Policiais olham para o Homem, que espera calmamente.

Policial 2: Hmm…

Policial 1: Quem é você?

Homem: Eu sou Daniel Dennett, seus idiotas.

Aparentemente do nada, Dennett puxa uma espada e corta a arma do Policial 2 ao meio com um som metálico. O Policial 1 começa a alcançar sua própria arma, mas Dennett está de repente atrás dele e golpeia com um punho, atingindo a junção do ombro e pescoço do Policial 1. O Policial 1 cai no chão.

O Policial 2 recua, horrorizado.

Policial 2: Isso não é possível! Como você fez isso?

Dennett: Eu sou um com meu corpo.

Dennett derruba o Policial 2 com outro golpe e avança em direção ao portão. Ele olha para o complexo de concreto imponente e segura sua espada com mais força.

Dennett *(falando consigo mesmo em voz baixa)*: Tem uma colher.

(Corte de volta para o General Fred e os outros oficiais militares.)

General Fred: Acabei de receber os relatórios. Perdemos Detroit.

Capitão Mudd: Não quero ser aquele que diz “Eu avisei”, mas -

General Fred: A Austrália foi… reduzida a átomos.

Coronel Todd: O vírus epifenomênico está se espalhando mais rápido. A própria civilização ameaça se dissolver em total normalidade. Podemos estar olhando para o fim da humanidade.

Capitão Mudd: Podemos negociar com os zumbis?

General Fred: Enviamos mensagens para eles. Eles enviaram apenas uma única resposta.

Capitão Mudd: Qual foi...?

General Fred: Está a caminho agora.

Um assistente traz um envelope e entrega ao General Fred.

General Fred abre o envelope, tira uma única folha de papel e lê.

O silêncio toma conta da sala.

Capitão Mudd: O que diz?

General Fred: Diz… que somos nós que estamos infectados com o vírus.

(Tudo fica silencioso)

O Coronel Todd ergue os braços e olha para suas mãos.

Coronel Todd: Meu Deus, é verdade. É verdade. Eu…

(Uma lágrima escorre pela bochecha do Coronel Todd.)

Coronel Todd: Eu não sinto nada.

A tela escurece.

O som fica silencioso.

O filme continua exatamente como antes.

227 - Excluindo o sobrenatural

Ocasionalmente, ouvimos alguém afirmando que o criacionismo não deveria ser ensinado nas escolas, especialmente não como uma hipótese concorrente à evolução, porque o criacionismo é *a priori* e automaticamente excluído da consideração científica, já que ele invoca o "sobrenatural".

Então... a ideia aqui é que o criacionismo poderia ser verdadeiro, mas mesmo se fosse verdadeiro, não seria permitido ensiná-lo na aula de ciências, porque a ciência trata apenas de coisas "naturais"?

Parece bastante claro que essa noção surge do desejo de evitar um confronto entre ciência e religião. Você não quer dizer diretamente que a ciência não ensina a *Afirmação Religiosa X* porque *X* foi testada pelo método científico e considerada falsa. Então, em vez disso, você pode... hum... alegar que a ciência está excluindo a hipótese *X* a priori. Assim, você não precisa discutir como o experimento desprovou a hipótese *X* a posteriori.

Claro, isso cai como uma luva na reivindicação criacionista de que o Projeto Inteligente não está recebendo um tratamento justo da ciência - que a ciência prejulgou a questão em favor do ateísmo, independentemente das evidências. Se a ciência excluísse o Projeto Inteligente a priori, esta seria uma queixa justificada!

Mas recuaremos um momento. Alguém vem até você e diz: "O Projeto Inteligente é excluído de ser ciência a priori, porque é 'sobrenatural', e a ciência lida apenas com explicações 'naturais'."

O que exatamente eles querem dizer com "sobrenatural"? Qualquer explicação inventada por alguém com o sobrenome "Cohen" é sobrenatural? Se vamos expulsar sumariamente um conjunto de hipóteses da ciência, o que exatamente devemos excluir?

De longe, a melhor definição que já ouvi do sobrenatural é a de Richard Carrier: Uma explicação "sobrenatural" apela para coisas mentais ontologicamente básicas, entidades mentais que não podem ser reduzidas a entidades não mentais.

Esta é a diferença, por exemplo, entre dizer que a água desce a colina porque quer estar mais baixa, e estabelecer equações diferenciais que afirmam descrever apenas movimentos, não desejos. É a diferença entre dizer que uma árvore produz folhas devido a um espírito arbóreo, versus examinar a bioquímica das plantas. A ciência cognitiva leva a luta contra o sobrenaturalismo para o reino da mente.

Por que esta é uma excelente definição do sobrenatural? Eu os remeto a Richard Carrier para o argumento completo. Mas considere: Suponha que você descubra o que parece um espírito, habitando uma árvore - uma dríade que pode se materializar fora ou na árvore, que fala em inglês sobre a necessidade de proteger sua árvore, e assim por diante. E então suponha que apontemos um microscópio para este espírito arbóreo, e ela se revele feita de partes - não partes inerentemente espirituais e inefáveis, como tecido de desejo e pano de crença, mas sim o mesmo tipo de partes que quarks e elétrons, partes cujo comportamento é definido em movimentos e não em mentes. A dríade não seria imediatamente rebaixada ao catálogo monótono das coisas comuns?

Mas se aceitarmos a definição de Richard Carrier do sobrenatural, surge um dilema: queremos dar às afirmações religiosas uma chance justa, mas parece que temos bons motivos para excluir explicações sobrenaturais a priori.

Quero dizer, como seria o universo se o reducionismo fosse falso?

Anteriormente, defini a tese reducionista assim: mentes humanas criam modelos de múltiplos níveis da realidade. Nesses modelos, padrões de alto nível e de baixo nível são representados separada e explicitamente. Um físico conhece a equação de Newton para a gravidade e a equação de Einstein para a gravidade. Ele também conhece a derivação da primeira como uma aproximação de baixa velocidade da segunda. Mas essas três representações mentais separadas são apenas uma conveniência da cognição humana. Não é que a própria realidade tenha uma equação de Einstein que governa em altas velocidades, uma equação de Newton que governa em baixas velocidades, e uma "lei de ligação" que suaviza a interface. A própria realidade tem apenas um único nível: a gravidade einsteiniana. É apenas a Falácia da Projeção Mental que faz algumas pessoas falarem como se os níveis mais altos pudessem ter uma existência separada - diferentes níveis de organização podem ter representações separadas nos mapas humanos, mas o território em si é um único objeto matemático unificado de baixo nível.

Suponha que isso estivesse errado.

Suponha que a Falácia da Projeção Mental não fosse uma falácia, mas simplesmente verdadeira.

Suponha que um Boeing 747 tivesse uma existência física fundamental separada dos quarks que o compõem.

Que observações experimentais você esperaria fazer, se se encontrasse em tal universo?

Se você não consegue dar uma boa resposta a isso, não é a observação que está descartando crenças "não reducionistas", mas uma incoerência lógica a priori. Se você não pode dizer quais previsões o modelo "não reducionista" faz, como pode dizer que a evidência experimental o descarta?

Minha tese é que o não reducionismo é uma confusão. E uma vez que você percebe que uma ideia é uma confusão, torna-se um pouco difícil imaginar como seria o universo se a confusão fosse verdadeira. Talvez eu tenha algum modelo de múltiplos níveis do mundo, e o modelo de múltiplos níveis tenha uma correspondência direta um-para-um com os elementos causais da física? Mas uma vez que todas as regras são especificadas, por que o modelo não se achataria em mais uma lista de coisas fundamentais e suas interações? Tudo que posso ver no modelo, como um 747 ou uma mente humana, tem que se tornar uma coisa real separada? Mas e se eu vir um padrão nesse novo supersistema?

O sobrenaturalismo é um caso especial de não reducionismo, onde não são os 747s que são irredutíveis, mas apenas (algumas) coisas mentais. A religião é um caso especial de sobrenaturalismo, onde as coisas mentais irredutíveis são Deus(es) e almas; e talvez também pecados, anjos, karma, etc.

Se eu propuser a existência de uma entidade poderosa capaz de examinar e alterar cada elemento do nosso universo observado, mas com essa entidade sendo redutível a partes não mentais que interagem com os elementos do nosso universo de maneira regular; se eu propuser que essa entidade deseja certas coisas específicas, mas "deseja" usando um cérebro composto de partículas e campos; então isso ainda não é uma religião, apenas uma hipótese naturalista sobre uma Matrix naturalista. Se amanhã as nuvens se abrissem e uma vasta figura amorfa e luminosa trovejasse a descrição acima da realidade, isso não implicaria que a figura fosse necessariamente honesta; mas eu exibiria os filmes em uma aula de ciências e tentaria derivar previsões testáveis da teoria.

Por outro lado, as religiões ignoraram a descoberta daquela antiga coisa sem corpo: onipresente no funcionamento da Natureza e imanente em cada folha que cai; vasta como a superfície de um planeta e com bilhões de anos; ela própria não criada e surgindo da estrutura da física; projetando sem cérebro para moldar toda a vida na Terra e as mentes da humanidade. A seleção natural, quando Darwin a propôs, não foi saudada como o tão aguardado Criador: ela não era fundamentalmente mental.

Mas agora chegamos ao dilema: se o entendimento convencional, monótono e normal da física e do cérebro estiver correto, não há maneira, em princípio, de um ser humano visualizar concretamente e derivar previsões experimentais testáveis sobre um universo alternativo no qual as coisas são irredutivamente mentais. Porque se o velho modelo normal e chato estiver correto, seu cérebro é feito de quarks, e assim seu

cérebro só conseguirá visualizar e prever concretamente coisas que podem ser previstas por quarks. Você só conseguirá construir modelos feitos de coisas simples interagindo.

Pessoas que vivem em universos reducionistas não podem visualizar concretamente universos não reducionistas. Elas podem pronunciar as sílabas "não reducionista", mas não podem imaginá-lo.

O erro básico do antropomorfismo, e a razão pela qual as explicações sobrenaturais parecem muito mais simples do que realmente são, é seu cérebro usando a si como uma caixa preta opaca para prever outras coisas rotuladas como "conscientes". Como você já tem grandes e complicadas redes de circuitos neurais que implementam seu "querer" coisas, parece que você pode facilmente descrever a água que "quer" fluir morro abaixo - a única palavra "querer" atua como uma alavanca para colocar sua própria maquinaria complexa de querer em movimento.

Ou você imagina que Deus goste de coisas bonitas e, portanto, fez as flores. Seu próprio circuito de "beleza" determina o que é "bonito" e "não bonito". Mas você não conhece o diagrama de suas próprias sinapses. Você não pode descrever um sistema não mental que calcula o mesmo rótulo para o que é "bonito" ou "não bonito" - não pode escrever um software que preveja seus próprios rótulos. Mas isso é apenas um defeito de conhecimento da sua parte; não significa que o cérebro não tenha explicação.

Se a "visão enfadonha" da realidade estiver correta, então você nunca poderá prever nada irredutível porque você é redutível. Você nunca poderá obter confirmação bayesiana para uma hipótese de irredutibilidade, porque qualquer previsão que você possa fazer é, portanto, algo que também poderia ser previsto por uma coisa redutível, a saber, seu cérebro.

Algumas situações não permitem pensar fora da caixa. Se o nosso universo for realmente computável por uma máquina de Turing, nunca conseguiremos imaginar concretamente algo que não seja computável por uma máquina de Turing — por mais que nossos matemáticos discutam vários níveis de hierarquias de oráculos de parada, não conseguiremos prever o que um oráculo de parada realmente diria, para distingui-lo experimentalmente de um raciocínio meramente computável.

Claro, isso tudo assumindo que a "visão enfadonha" está correta. Na medida em que você acredita que a evolução é verdadeira, você não deve esperar encontrar evidências fortes contra a evolução. Na medida em que você acredita que o reducionismo é verdadeiro, você deve esperar que hipóteses não reducionistas sejam incoerentes além de erradas. Na medida em que você acredita que o sobrenaturalismo é falso, você deve esperar que ele seja inconcebível também.

Se, por outro lado, uma hipótese sobrenatural se revelar verdadeira, então presumivelmente você também descobrirá que ela não é inconcebível.

Então, traremos isso de volta ao círculo completo para a questão do Projeto Inteligente:

O PI deve ser excluído a priori da falsificação experimental e das salas de aula de ciências porque, ao invocar o sobrenatural, ele se colocou fora da filosofia natural?

Respondo: "Claro que não". A irredutibilidade do projetista inteligente não é uma parte indispensável da hipótese do PI. Para cada Deus irredutível que pode ser proposto pelos defensores do PI, existe um alienígena redutível correspondente que se comporta segundo as mesmas previsões, já que os próprios defensores do PI são redutíveis. Na medida em que acredito que o reducionismo está de fato correto, o que é uma medida bastante forte, devo esperar descobrir formulações redutíveis de todos os modelos preditivos supostamente sobrenaturais.

Se estamos examinando os registros arqueológicos para testar a afirmação de que Jeová dividiu o Mar Vermelho por um desejo explícito de exibir seu poder sobre-humano, então faz pouca diferença se Jeová é ontologicamente básico, ou um alienígena com nanotecnologia, ou um Senhor das Trevas da Matrix. Você faz alguma arqueologia, não encontra restos de esqueletos ou armaduras no local do Mar Vermelho, e de fato encontra registros de que o Egito governava grande parte de Canaã na época. Então você carimba o registro histórico na Bíblia como "refutado" e segue em frente. A hipótese é coerente, falsificável e errada.

O mesmo acontece com as evidências da biologia de que as raposas são projetadas para perseguir coelhos, os coelhos são projetados para fugir das raposas, e nenhum deles é projetado "para continuar sua espécie" ou "proteger a harmonia da Natureza"; o mesmo acontece com a retina sendo projetada ao contrário, com as partes sensíveis à luz na parte inferior; e assim por diante, através de milhares de outros itens de evidência para um projeto fragmentado, imoral e incompetente. O modelo Jeová do nosso deus alienígena é coerente, falsificável e errado - coerente, isto é, desde que você não se importe se Jeová é ontologicamente básico ou apenas um alienígena.

Basta converter a hipótese sobrenatural na hipótese natural correspondente. Basta fazer as mesmas previsões da mesma maneira, sem afirmar que coisas mentais são ontologicamente básicas. Consulte a caixa preta do seu cérebro se necessário para fazer previsões - por exemplo, se você quiser falar sobre um "deus zangado" sem construir uma IA zangada completa para rotular comportamentos como zangados ou não zangados. Assim, você deriva as previsões, ou procura as previsões feitas por antigos teólogos sem conhecimento prévio dos nossos resultados experimentais. Se o experimento entrar em conflito com essas previsões, então é justo falar que a afirmação religiosa foi cientificamente refutada. Foi dada sua justa chance de confirmação; está sendo excluída a posteriori, não a priori.

Em última análise, o reducionismo é apenas descrença em coisas fundamentalmente complicadas. Se "fundamentalmente complicado" soa como um oximoro, uma contradição em termos… bem, é por isso que penso que a doutrina do não reducionismo é uma confusão, em vez de uma maneira como as coisas poderiam ser, mas não são. Você seria sábio em ser cauteloso, se você se encontrar supondo tais coisas.

Mas a regra última da ciência é olhar e ver. Se alguma vez um Deus aparecesse para trovejar sobre as montanhas, seria algo que as pessoas olhariam e veriam.

Corolário: Qualquer suposto projetista de Inteligência Artificial Geral que fale sobre crenças religiosas em tons respeitosos claramente não é um especialista em reduzir coisas mentais a coisas não mentais; e de fato sabe tão pouco dos fundamentos mais básicos, a ponto de ser pouco plausível que possam ser especialistas na arte; a menos que o savantismo deles seja completo. Ou, é claro, se estiverem mentindo descaradamente. Não estamos falando de um erro sutil.

228 - Poderes psíquicos

No último ensaio, eu escrevi:

Se a "visão entediante" da realidade estiver correta, então você nunca poderá prever algo irredutível porque você é redutível. Você nunca pode obter confirmação Bayesiana para uma hipótese de irredutibilidade, porque qualquer previsão que você faça é, portanto, algo que também poderia ser previsto por algo redutível, a saber, seu cérebro.

Benja Fallenstein comentou:

> Eu acho que, embora você não possa, neste caso, nunca criar um teste empírico cujo resultado possa logicamente provar irredutibilidade, não há uma razão clara para acreditar que você não possa criar um teste cujo resultado contrafactual em um mundo irredutível tornaria a irredutibilidade subjetivamente muito mais provável (dada uma prioridade occamiana).
>
> Sem entrar na questão da redutibilidade/irredutibilidade, considere o cenário no qual o universo físico possibilita construir um hiper-computador - que realiza operações com números reais arbitrários, por exemplo - mas que nossos cérebros, na verdade, não fazem uso disso: eles podem ser simulados perfeitamente bem por uma máquina de Turing comum, muito obrigado…

Bem, isso é um argumento muito inteligente, Benja Fallenstein. Mas tenho uma resposta esmagadora para seu argumento, tal que, uma vez que eu a der, você desistirá imediatamente de continuar debatendo comigo sobre este ponto em particular:

Você está certo.

Infelizmente, não ganho pontos de modéstia desta vez, porque após publicar o último ensaio, percebi uma falha semelhante por conta própria sobre a Navalha de Occam e poderes psíquicos:

Se crenças e desejos são entidades ontologicamente básicas e irredutíveis, ou têm um componente ontologicamente básico não coberto pela ciência existente, isso tornaria muito mais provável que houvesse uma regra ontológica governando a interação de diferentes mentes—uma interação que evitava meios de comunicação "materiais" ordinários como ondas sonoras, conhecidas pela ciência existente.

Se o naturalismo estiver correto, então existe um modelo reducionista conjugado que faz as mesmas previsões que qualquer previsão concreta que qualquer parapsicólogo possa fazer sobre telepatia.

De fato, se o naturalismo estiver correto, a única razão pela qual podemos conceber crenças como "fundamentais" é devido à falta de autoconhecimento de nossos próprios neurônios - que a arquitetura reflexiva peculiar de nossas próprias mentes expõe a classe "crença", mas oculta a maquinaria por trás dela.

Mesmo assim, a descoberta de transferência de informações entre cérebros, na ausência de qualquer conexão material conhecida entre eles, é probabilisticamente uma previsão privilegiada de modelos sobrenaturais (aqueles que contêm entidades mentais ontologicamente básicas). Isso porque é muito mais simples nesse caso ter uma nova lei relacionando crenças entre diferentes mentes, comparado ao modelo "entediante" onde crenças são construções complexas de neurônios.

A esperança de poderes psíquicos surge do tratamento de crenças e desejos como objetos suficientemente fundamentais que podem ter conexões não mediadas com a realidade. Se as crenças são apenas padrões de neurônios feitos de coisas que conhecemos, com entradas dadas por órgãos como os olhos, feitos de material conhecido, e com saídas através de músculos feitos de material conhecido, e isso parece o suficiente para explicar tudo o que sabemos sobre a mente humana, não tem por que pensar que há algo mais além disso - não há motivo para postular conexões adicionais. É por isso que os reducionistas não acreditam em poderes psíquicos. Se alguém realmente visse um poder psíquico, isso seria uma prova bem forte de que existe algo sobrenatural, como o Richard Carrier diz.

Temos uma regra de Occam que conta o número de classes ontologicamente básicas e leis ontologicamente básicas no modelo, e penaliza a contagem de entidades. Se o naturalismo estiver correto, então a tentativa de contar "crença" ou a "relação entre crença e realidade" como uma única entidade básica é simplesmente um antropomorfismo equivocado; somos apenas tentados a isso por uma peculiaridade da arquitetura interna do nosso cérebro. Mas se você seguir essa visão equivocada, então ela atribui uma probabilidade muito maior a poderes psíquicos do que o naturalismo, porque você pode implementar poderes psíquicos usando leis aparentemente mais simples.

Por isso, a descoberta real de poderes psíquicos implicaria que a regra de Ocam ingênua humana estava, de fato, melhor calibrada do que a sofisticada regra de Ocam naturalista. Isso argumentaria que os reducionistas estavam errados o tempo todo em tentar dissecar o cérebro; que o que nossas mentes expunham como uma alavanca aparentemente simples era, de fato, uma alavanca simples. Os dualistas ingênuos teriam estado certos desde o início, o que é por isso que seu antigo desejo teria sido habilitado para se tornar realidade.

Portanto, a telepatia, a habilidade de influenciar eventos apenas os desejando, e a precognição, se descobertos, seriam fortes evidências Bayesianas a favor da hipótese de que crenças são ontologicamente fundamentais. Não uma prova lógica, mas fortes evidências Bayesianas.

Se o reducionismo estiver correto, então qualquer história de ficção científica contendo poderes psíquicos pode ser produzida por um sistema de elementos simples (isto é, o cérebro do autor da história); mas se de fato descobrirmos poderes psíquicos, isso tornaria muito mais provável que eventos estavam ocorrendo que não poderiam, de fato, ser descritos por modelos reducionistas.

Isso só quer dizer: a existência de poderes psíquicos é uma afirmação probabilística privilegiada de visões de mundo não-reducionistas - eles possuem essa previsão antecipada; eles a conceberam e a propuseram, em desafio às expectativas reducionistas.

Então, pelas leis da ciência, se poderes psíquicos forem descobertos, o não-reducionismo vence.

Portanto, estou confiante em descartar poderes psíquicos como a priori implausíveis, apesar de todas as evidências experimentais reivindicadas a favor deles.

Parte S - Física Quântica e muitos mundos

229 - Explicações quânticas

Existe uma crença generalizada de que a mecânica quântica supostamente é confusa. Essa não é uma boa mentalidade, nem para um professor, nem para um aluno.

Percebo que assuntos lendariamente tidos como "confusos" frequentemente não são tão complicados matematicamente, especialmente se você busca apenas uma compreensão básica - mas ainda matemática - do que acontece lá no fundo.

Não sou físico, e os físicos são famosos por detestar quando não profissionais falam sobre mecânica quântica. Mas tenho alguma experiência em explicar coisas matemáticas supostamente "difíceis de entender".

Escrevi a "Explicação Intuitiva do Raciocínio Bayesiano" porque as pessoas reclamavam que o Teorema de Bayes era "contra-intuitivo" - na verdade, era notoriamente contra-intuitivo - e isso não me parecia correto. A equação simplesmente não parecia complexa o suficiente para merecer a temível reputação que tinha. Então, tentei explicá-la à minha maneira. Não consegui atingir meu objetivo original de alcançar alunos do ensino fundamental, mas recebo frequentes e-mails de agradecimento de pessoas antes confusas, desde jornalistas até professores universitários de outras áreas.

Além disso, como bayesiano, não acredito em fenômenos inerentemente confusos. A confusão existe em nossos modelos do mundo, não no próprio mundo. Se um assunto é amplamente conhecido como confuso, não apenas difícil… você não deveria se contentar com isso. Não é satisfatório; não é um lugar aceitável para estar. Talvez você possa resolver o problema, talvez não; mas não deveria ficar satisfeito em deixar os alunos confusos.

A primeira maneira pela qual minha introdução vai se afastar da introdução tradicional e padrão à mecânica quântica é que não vou lhe dizer que a mecânica quântica deve ser confusa.

Não vou lhe dizer que tudo bem você não entender mecânica quântica, porque ninguém entende mecânica quântica, como Richard Feynman uma vez alegou. Houve um período histórico em que isso era verdade, mas não vivemos mais nessa era.

Não vou lhe dizer: "Você não entende mecânica quântica, você apenas se acostuma com ela." (Como von Neumann supostamente disse; lá nas décadas sombrias, quando, de fato, ninguém entendia mecânica quântica.)

Explicações devem deixá-lo menos confuso. Se você sente que não entende alguma coisa, isso indica um problema - seja com você ou com seu professor - mas de qualquer forma, um problema; e você deve agir para resolvê-lo.

Não vou lhe dizer que a mecânica quântica é estranha, bizarra, confusa ou alienígena. A mecânica quântica é contra-intuitiva, mas isso é um problema com suas intuições, não um problema com a mecânica quântica. A mecânica quântica existia há bilhões de anos antes de o Sol se condensar do hidrogênio interestelar. A mecânica quântica estava aqui antes de você, e se você tem um problema com isso, é você quem precisa mudar. A mecânica quântica certamente não mudará. Não existem fatos surpreendentes, apenas modelos surpreendidos por fatos; e se um modelo é surpreendido pelos fatos, isso não é mérito desse modelo.

É sempre melhor pensar na realidade como perfeitamente normal. Desde o início, nenhuma coisa incomum jamais aconteceu.

O objetivo é sentir-se completamente à vontade em um universo quântico. Como um nativo. Porque, de fato, é aí que você vive.

Na sequência que virá sobre mecânica quântica, vou consistentemente falar como se a mecânica quântica fosse perfeitamente normal; e quando as intuições humanas se afastarem da mecânica quântica, zombarei das intuições por serem estranhas e incomuns. Isso pode parecer estranho, mas o ponto é fazer sua mente girar para um ponto de vista quântico nativo.

Outra coisa: A introdução tradicional à mecânica quântica segue de perto a ordem na qual a mecânica quântica foi descoberta.

A introdução tradicional começa dizendo que a matéria às vezes se comporta como pequenas bolas de bilhar saltitantes, e às vezes se comporta como picos e vales movendo-se através de uma piscina de água. Então, a introdução tradicional dá alguns exemplos da matéria agindo como uma pequena bola de bilhar, e alguns exemplos dela agindo como uma onda oceânica.

Agora, acontece que é um fato histórico que, quando os estudiosos da matéria estavam descobrindo tudo isso e não tinham ideia sobre a verdadeira matemática subjacente, aqueles primeiros cientistas inicialmente pensaram que a matéria era como pequenas bolas de bilhar. E depois que era como ondas no oceano. E então que era como bolas de bilhar novamente. E então os primeiros cientistas ficaram realmente confusos e permaneceram assim por várias décadas, até que finalmente tudo foi esclarecido na segunda metade do século XX.

Arrastar um estudante moderno por tudo isso pode ser uma abordagem historicamente realista do assunto, mas também garante o resultado historicamente realista de total perplexidade. Falar com jovens físicos aspirantes sobre “dualidade onda/partícula” é como iniciar os estudantes de química pelos Quatro Elementos.

Um elétron não é uma bola de bilhar, e não é uma crista e um vale movendo-se por meio de uma piscina de água. Um elétron é uma entidade matematicamente diferente, o tempo todo e em todas as circunstâncias, e deve ser aceito em seus próprios termos.

O universo não está oscilando entre usar partículas e ondas, incapaz de se decidir. São apenas as intuições humanas sobre mecânica quântica que alternam. As intuições que temos para bolas de bilhar e as intuições que temos para cristas e vales em uma piscina de água, ambas parecem de alguma forma aplicáveis aos elétrons, em momentos diferentes e sob circunstâncias diferentes. Mas a verdade é que ambas as intuições simplesmente não são aplicáveis.

Se você tentar pensar em um elétron como se ele fosse parecido com uma bola de bilhar em alguns dias, e como uma onda oceânica em outros dias, você vai se confundir completamente.

No entanto, são seus olhos que estão oscilando e instáveis, não o mundo.

Além disso:

A ordem na qual a humanidade descobriu as coisas não é necessariamente a melhor ordem para ensiná-las. Primeiro, a humanidade notou que havia outros animais correndo por aí. Então nós os abrimos e descobrimos que eles estavam cheios de órgãos. Então examinamos os órgãos cuidadosamente e descobrimos serem feitos de tecidos. Então olhamos para os tecidos sob um microscópio e descobrimos células, as quais são feitas de proteínas e algumas outras coisas sintetizadas quimicamente. Que são feitas de moléculas, feitas de átomos, feitos de prótons e nêutrons e elétrons, que são muito mais simples do que animais inteiros, mas foram descobertos dezenas de milhares de anos depois.

A física não começa falando sobre biologia. Então, por que deveria começar falando sobre fenômenos muito complicados de alto nível, como, digamos, os resultados observados de experimentos?

A maneira comum de ensinar mecânica quântica continua enfatizando os resultados experimentais. Agora, eu realmente entendo por que isso soa bom de uma perspectiva racionalista. Acredite em mim, eu entendo.

Mas me parece que o resultado é arrastar grandes ferramentas matemáticas complicadas que você precisa para analisar situações reais, antes que o aluno entenda o que fundamentalmente acontece nos casos mais simples.

É como tentar ensinar programadores a escrever programas concorrentes multithreaded antes que eles saibam como somar duas variáveis, porque programas concorrentes multithreaded estão mais próximos da vida cotidiana. Estar próximo da vida cotidiana nem sempre é uma forte recomendação para o que ensinar primeiro.

Talvez o foco monomaníaco nas observações experimentais fizesse sentido nas décadas sombrias, quando ninguém entendia o que estava fundamentalmente acontecendo, e você não podia começar por aí, e todos os seus modelos eram apenas matemáticas misteriosas que davam boas previsões experimentais… você ainda pode encontrar essa visão da física quântica apresentada em muitos livros… mas talvez hoje valha a pena tentar um ângulo diferente? O resultado da abordagem padrão é a confusão padrão.

O mundo clássico está estritamente implícito no mundo quântico, mas ver de uma perspectiva clássica torna tudo maior e mais complicado.

A vida cotidiana é um nível mais alto de organização, como moléculas versus quarks - enorme catálogo de moléculas, seis quarks. Acho que vale a pena tentar ensinar primeiro da perspectiva do mundo quântico e falar sobre resultados experimentais clássicos depois.

Não começarei com o mundo clássico normal e depois falar sobre um pano de fundo quântico bizarro escondido nos bastidores. O mundo quântico é o cenário e ele define a normalidade.

Não falarei como se o mundo clássico fosse a realidade, e ocasionalmente o mundo clássico transmitisse um pedido de resultado experimental para um servidor de física quântica, e o servidor de física quântica fizesse alguns cálculos peculiares e transmitisse de volta um resultado experimental clássico. Falarei como se o mundo quântico fosse o verdadeiramente real e o mundo clássico algo distante. Não apenas porque isso facilita ser um nativo de um universo quântico, mas porque, em um nível fundamental, é a verdade.

Finalmente, vou adotar uma perspectiva estritamente realista sobre a mecânica quântica - o mundo quântico está realmente lá fora, nossas equações descrevem o território e não nossos mapas dele, e o mundo clássico só existe implicitamente no quântico. Não discutirei visões não realistas nos estágios iniciais da minha introdução, exceto para dizer por que você não deve se confundir com certas intuições que os não realistas usam como suporte. Não vou me desculpar por isso, e gostaria de pedir a quaisquer não realistas sobre o assunto da mecânica quântica que esperem e segurem seus comentários até serem solicitados em um ensaio posterior. Façam-me esse favor, por favor. Acho que o não realismo é uma das principais coisas que confunde osestudantes em potencial e os impede de visualizar concretamente os fenômenos quânticos. Discutirei as questões explicitamente em um ensaio futuro.

Mas todos devem estar cientes de que, mesmo que eu não vá discutir o assunto no início, existe uma comunidade considerável de cientistas que contestam a perspectiva realista da mecânica quântica. Eu, pessoalmente, não acho que vale a pena considerar as duas formas; sou um realista puro, por razões que se tornarão aparentes. Mas se você ler minha introdução, você está obtendo minha visão. Não é apenas a minha visão. É provavelmente a visão majoritária entre os físicos teóricos, se isso conta para alguma coisa (embora eu vá argumentar a questão separadamente das pesquisas de opinião). Ainda assim, não é a única visão que existe na comunidade física moderna. Não me sinto obrigado a apresentar as outras visões imediatamente, mas me sinto obrigado a avisar meus leitores que existem outras visões, que não apresentarei durante os estágios iniciais da introdução.

Para resumir, meu objetivo será ensinar você a pensar como um nativo de um universo quântico, não como um turista relutante.

Abrace a realidade. Abrace-a com força.

230 - Configurações e amplitude

Então, o universo não é composto de pequenas bolas de bilhar, nem de picos e vales em um éter infinito. Então, de que coisa é feita esta coisa?

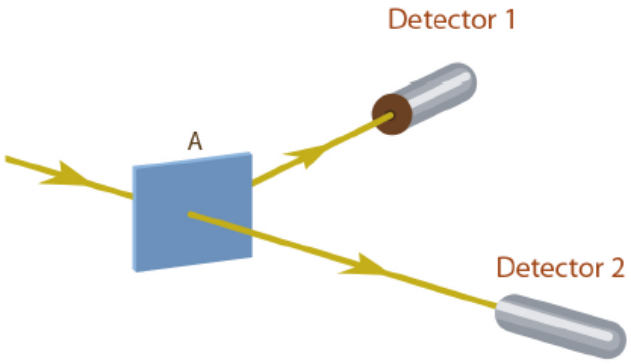


Figura 230.1

Na Figura 230.1, podemos observar em A um espelho semi-refletor e dois detectores de fótons, Detector 1 e Detector 2.

Os primeiros cientistas, ao realizar experimentos como este, ficaram perplexos com o significado dos resultados. Ao enviar um fóton em direção ao espelho semi-refletor, metade das vezes observavam o Detector 1 clicar, e a outra metade das vezes, o Detector 2 clicar.

Os primeiros cientistas - é de dar risada - pensavam que o espelho semi-refletor desviava o fóton metade das vezes e o deixava passar a outra metade.

Ha, ha! Como se o espelho semi-refletor agisse de maneira diferente em ocasiões diferentes! Gostaria que você abandonasse essa ideia, pois, se apegar ao que os primeiros cientistas pensavam só resultará em extrema confusão. O espelho semi-refletor sempre obedece à mesma regra.

Se você fosse escrever um software que simulasse esse experimento - não um programa que previsse o resultado do experimento, mas um programa que se assemelhasse à realidade subjacente - ele poderia ser mais ou menos assim:

No início do programa (o início do experimento, o início do tempo), existe uma entidade matemática chamada de “configuração.” Pode-se pensar nessa configuração como algo que representa “um fóton vindo da fonte de fótons em direção ao espelho semi-refletor” ou simplesmente “um fóton se dirigindo a A”.

Uma configuração pode armazenar um único valor complexo - “complexo” como nos números complexos (a + bi), onde i é definido como . No início do programa, já existe um número complexo armazenado na configuração “um fóton indo em direção de “A”. O valor exato não importa, desde que não seja zero. Atribuiremos o valor (−1 + 0i) à configuração “um fóton indo em direção de “A”.

Tudo isso é um fato no domínio do programa, não uma descrição do conhecimento de alguém. Uma configuração não é uma proposição nem uma possível maneira de o mundo ser. Uma configuração é uma variável no programa - você pode pensá-la como um tipo de local de memória cujo índice é “um fóton indo em direção de “A” - e ela está lá fora, no território.

Como os números complexos atribuídos às configurações não são números reais positivos entre 0 e 1, não há risco de confundi-los com probabilidades. A configuração “um fóton indo em direção de “A” tem valor complexo -1, o que é difícil de ver como um grau de confiança. Os números complexos são valores dentro do programa, mais uma vez, lá fora, no território. Chamaremos esses números complexos de amplitudes.

Existem mais duas configurações, que denominaremos “um fóton indo de A para o Detector 1” e “um fóton indo de A para o Detector 2”. Essas configurações ainda não possuem um valor complexo; ele é atribuído à medida que o programa é executado.

Vamos calcular as amplitudes de “um fóton indo de A para 1” e “um fóton indo de A para 2” usando o valor de “um fóton indo para A” e a regra que descreve o espelho semi-refletor em A.

Em termos gerais, a regra do espelho semi-refletor é a seguinte: “multiplique por 1 quando o fóton seguir em linha reta e multiplique por i quando o fóton girar em um ângulo reto”. Essa é a regra universal que relaciona a amplitude da configuração “um fóton entrando” com a amplitude que vai para as configurações de “um fóton saindo em linha reta” ou “um fóton sendo desviado”. [1]

Portanto, encaminhamos a amplitude da configuração “um fóton indo em direção a A,” que é (−1 + 0i), para o espelho semi-refletor em A, e isso transmite uma amplitude de (−1 + 0i) × i = (0 − i) para “um fóton indo de A para 1” e também transmite uma amplitude de (−1 + 0i) × 1 = (−1 + 0i) para “um fóton indo de A para 2”.

No experimento da Figura 230.1, essas são todas as configurações e amplitudes transmitidas com as quais precisamos nos preocupar, então concluímos. Ou, se preferir pensar em “Detector 1 recebe um fóton” e “Detector 2 recebe um fóton” como configurações separadas, elas apenas herdam seus valores de “A para 1” e “A para 2,” respectivamente. (Na realidade, os valores herdados devem ser multiplicados por outro fator complexo correspondente à distância de “A” até o detector; mas ignoraremos isso por enquanto e supor que todas as distâncias percorridas em nossos experimentos correspondam a um fator complexo de 1.)

Assim, o estado final do programa é:

Configuração “um fóton indo em direção a A”: (−1 + 0i)

Configuração “um fóton indo de A para 1”: (0 − i)

Configuração “um fóton indo de A para 2”: (−1 + 0i)

E, opcionalmente,

Configuração “Detector 1 recebe um fóton”: (0 − i)

Configuração “Detector 2 recebe um fóton”: (−1 + 0i).

Este mesmo resultado ocorre - as mesmas amplitudes armazenadas nas mesmas configurações - toda vez que você executa o programa (ou seja, toda vez que realiza o experimento).

As razões pelas quais não podemos medir diretamente as amplitudes exatas de cada configuração são complexas e extrapolam o escopo da mecânica quântica fundamental. Elas pertencem a um nível de organização mais alto, assim como os átomos são mais complexos que os quarks. Por conseguinte, não existe

um instrumento de medição simples capaz de determinar diretamente o estado do programa.

Então, como os físicos sabem quais são as amplitudes?

Temos uma ferramenta de medição mágica que pode nos dizer o módulo ao quadrado da amplitude de uma configuração. Se a amplitude complexa original for (a + bi), podemos obter o número real positivo ($a^2 + b^2$). Pense no teorema de Pitágoras: se você imaginar o número complexo como uma pequena seta que se estende desde a origem em um plano bidimensional, a ferramenta mágica nos diz o comprimento ao quadrado da pequena seta, mas não nos diz a direção para a qual a seta aponta.

Para ser mais preciso, a ferramenta mágica, na verdade, apenas nos fornece as proporções dos comprimentos quadrados das amplitudes em algumas configurações. Não sabemos quanto tempo as setas têm em termos absolutos, apenas quanto tempo elas têm em relação umas às outras. Mas isso acaba sendo informação suficiente para nos permitir reconstruir as leis da física - as regras do programa. E assim, posso falar sobre amplitudes, não apenas proporções de módulos quadrados.

Quando usamos a ferramenta mágica para “Detector 1 obtém um fóton” e “Detector 2 obtém um fóton”, descobrimos que essas configurações têm o mesmo módulo ao quadrado - os comprimentos das setas são os mesmos. Assim diz a ferramenta mágica. Realizando experimentos mais complexos (que veremos em breve), podemos determinar que os números complexos originais tinham uma relação de i para 1.

E o que é essa ferramenta de medição mágica?

Na vida cotidiana, que fica muito acima do nível quântico e é muito mais complexa, a ferramenta de medição mágica funciona enviando fótons para um espelho semi-refletor, um de cada vez, e contando quantos fótons chegam ao Detector 1 em comparação com o Detector 2 em alguns milhares de tentativas. A razão entre esses valores corresponde à razão dos módulos quadrados das amplitudes. Mas a razão para isso é algo que ainda não exploramos. Devemos aprender a andar antes de correr. Não é possível compreender o que acontece no nível da vida cotidiana antes de entender o que acontece em casos muito mais simples.

Para os fins de hoje, dispomos de um leitor mágico de razão de módulo quadrado. E a ferramenta mágica nos informa que o comprimento quadrado da pequena seta bidimensional para a configuração “Detector 1 obtém um fóton” é igual ao comprimento quadrado para “Detector 2 obtém um fóton.” É isso.

Você pode se perguntar: “Dado que a ferramenta mágica funciona dessa maneira, o que nos motiva a utilizar a teoria quântica em vez de acreditar que o espelho semi-refletor reflete o fóton cerca da metade do tempo?”

Bem, isso só serviria para nos confundir - adotar uma mentalidade historicamente realista como essa e utilizar intuições cotidianas. Lembra-se de eu mencionar alguma coisa sobre uma bolinha de bilhar indo de um lado para o outro e talvez quicando em um espelho? Não é assim que a realidade funciona. A realidade envolve amplitudes complexas fluindo entre configurações, e as leis desse fluxo são constantes.

Mas, se você insiste em imaginar uma situação mais complexa que as formas de pensar com bolas de bilhar não conseguem compreender, aqui está um experimento mais complicado.

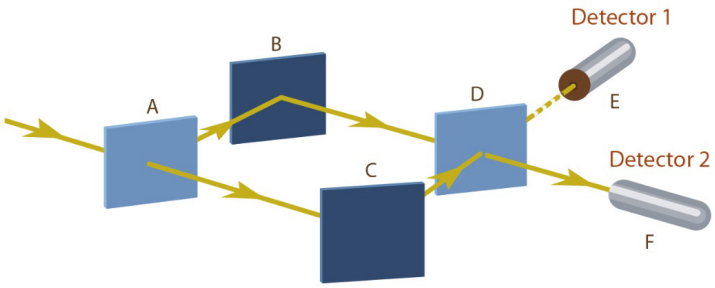


Figura 230.2

Na Figura 230.2, B e C são espelhos completos, enquanto A e D são semi-espelhos. A linha de “D” para “E” é tracejada por razões que se tornarão evidentes, mas a amplitude está fluindo de “D” para “E” sob as mesmas leis.

Agora, aplicaremos as regras que aprendemos anteriormente:

No início do tempo, “um fóton indo em direção a A” tem a amplitude (−1 + 0*i*).”

Calculamos agora as amplitudes para as configurações ‘um fóton indo de A para B’ e ‘um fóton indo de A para C’:

”Um fóton indo de A para B” = *i* × ‘um fóton indo em direção a A” = (0 − *i*).

Da mesma forma,

“Um fóton indo de A para C” = 1 × ”um fóton indo em direção a A” = (−1 + 0*i*).

Os espelhos completos se comportam (como seria de esperar) como metade de um espelho semi-refletor - um espelho completo apenas dobra a luz em ângulos retos e a multiplica por i. (Para ser um pouco mais preciso: para um espelho completo, a amplitude que flui da configuração de um fóton entrando para a configuração de um fóton saindo em um ângulo reto é multiplicada por um fator de i.)

Portanto:

“Um fóton indo de B para D” = i × “um fóton indo de A para B” = (1 + 0i).

“Um fóton indo de C para D” = i × “um fóton indo de A para C” = (0 − i).

‘Fóton de B para D’ e ‘fóton de C para D’ são duas configurações diferentes – não escrevemos simplesmente ‘um fóton em D’ – porque os fótons estão chegando em dois ângulos diferentes nessas duas configurações diferentes. E o que D faz com um fóton depende do ângulo no qual o fóton chega.

Novamente, a regra (falando livremente) é que quando um espelho semi-refletor curva a luz em um ângulo reto, a amplitude que flui da configuração de entrada do fóton para a configuração de saída do fóton é a amplitude do fóton - configuração inicial multiplicada por i. E quando duas configurações são relacionadas por meio de um espelho que deixa a luz passar, a amplitude que flui da configuração de entrada do fóton é multiplicada por 1.

Portanto:

- Da configuração “um fóton indo de B para D”, com amplitude original (1 + 0i):
- Amplitude de (1 + 0i) × i = (0 + i) flui para “um fóton indo de D para E”.
- Amplitude de (1 + 0i) × 1 = (1 + 0i) flui para “um fóton indo de D para F”.

- Da configuração “um fóton indo de C a D”, com amplitude original (0 − i):
- Amplitude de (0 − i) × i = (1 + 0i) flui para “um fóton indo de D para F”.
- Amplitude de (0 − i) × 1 = (0 − i) flui para “um fóton indo de D para E”.

Portanto:

- A amplitude total fluindo para a configuração “um fóton indo de D para E” é (0 + i) + (0 − i) = (0 + 0i) = 0.
- A amplitude total fluindo para a configuração “um fóton indo de D a F” é (1 + 0i) + (1 + 0i) = (2 + 0i).

(Você pode tentar resolver isso sozinho com papel e caneta se perder a linha de raciocínio em algum momento.)

No entanto, o resultado, dessa perspectiva “experimental” de alto nível que consideramos como vida normal, é que não observamos fótons detectados em E. Todos os fótons parecem terminar em F. A razão dos módulos quadrados entre ‘D para E’ e ‘D para F’ é de 0 a 4. É por isso que a linha de “D” para “E” é tracejada nesta figura.

Isso não pode ser explicado pensando em espelhos semi-refletores desviando pequenas bolas de bilhar em metade do tempo. Você deve pensar em termos de fluxos de amplitude.

Se os espelhos semi-refletores desviassem uma bola de bilhar na metade das vezes, nessa configuração, a bola acabaria no Detector 1 na metade das vezes e no Detector 2 na metade das vezes. Isso não acontece. Portanto, não considere essa possibilidade.

Você pode argumentar: “Mas espere um minuto! Posso pensar em outra hipótese que explique esse resultado. E se, quando um espelho semi-refletor reflete um fóton, ele faz algo com o fóton para garantir que ele não seja refletido na próxima vez? E, quando deixa um fóton passar diretamente, ele faz algo com o fóton para que ele seja refletido na próxima vez.”

Na realidade, não há necessidade de complicar as regras. Lembre-se da Navalha de Occam. Basta usar fluxos simples de amplitude normal entre as configurações.

Mas se desejar outro experimento que refute sua nova hipótese alternativa, veja a Figura 230.3.

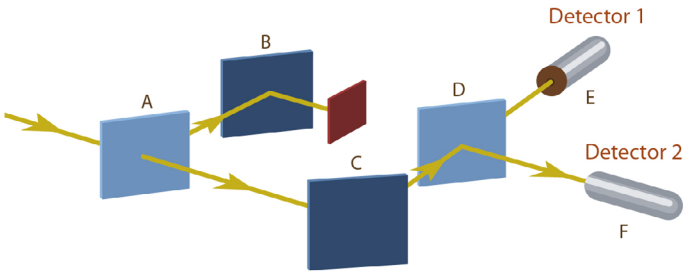


Figura 230.3

Aqui, mantemos todas as configurações experimentais iguais e apenas inserimos um pequeno objeto de bloqueio entre B e D. Isso assegura que a amplitude de ‘um fóton indo de B para D’ seja 0.

Após eliminar as contribuições de amplitude dessa configuração, você fica com totais de (1 + 0i) em ‘um fóton indo de D para F’ e (0 - i) em ‘um fóton indo de D para E’.

Os módulos quadrados de (1 + 0i) e (0 - i) são ambos 1, então a ferramenta de medição mágica deve nos dizer que a razão dos módulos quadrados é 1. Retornando ao nível onde os físicos existem, descobrimos que o Detector 1 desliga metade do tempo e o Detector 2 desliga metade do tempo.

A mesma coisa acontece se colocarmos o bloqueio entre C e D. As amplitudes são diferentes, mas a razão dos módulos quadrados ainda é 1, então o Detector 1 desliga metade do tempo e o Detector 2 desliga metade do tempo.

Isso não pode ocorrer com uma pequena bola de bilhar que é refletida ou não pelos espelhos semi-refletores.

Devido aos números complexos terem direções opostas, como 1 e -1, ou i e -i, os fluxos de amplitude podem se anular. A amplitude fluindo da configuração X para a configuração Y pode ser cancelada por uma amplitude igual e oposta fluindo da configuração Z para a configuração Y., na verdade, isso é exatamente o que acontece neste experimento.

Na teoria da probabilidade, quando algo pode acontecer de uma maneira ou de outra, X ou ¬X, então P(Z) = P(Z|X)P(X) + P(Z|¬X)P(¬X). E todas as probabilidades são positivas. Então, se você estabelecer que a probabilidade de Z acontecer, dado que X aconteceu, é ½, e a probabilidade de X acontecer é 1/3. Portanto, a probabilidade total de Z acontecer é de pelo menos 1/6, independentemente do que aconteça no caso de ¬X. Não existem probabilidades negativas, credibilidade menor que impossível ou credibilidade (0 + i). Consequentemente, os graus de crença não podem se anular como as amplitudes.

Além disso, vale lembrar que, para começar, a probabilidade está na mente, e estamos falando do território, do programa que é a realidade, não da cognição humana ou dos estados de conhecimento parcial.

Da mesma forma, as configurações não são proposições, afirmações ou modos como o mundo poderia ser. As configurações não são construções semânticas. Adjetivos como 'provável' não se aplicam a elas; não são crenças, sentenças ou mundos possíveis. Elas não são verdadeiras ou falsas, mas simplesmente reais.

No experimento da Figura 230.2, evite pensar algo como: "O fóton vai para B ou C, mas poderia ter ido para o outro lado, e essa possibilidade interfere em sua capacidade de ir para E..."

Não faz sentido pensar que algo que "poderia ter acontecido, mas não aconteceu" exerce um efeito sobre o mundo. Podemos imaginar coisas que poderiam ter acontecido, mas não aconteceram, como pensar: "Puxa, aquele carro quase me atropelou", e nossa imaginação pode afetar nosso comportamento futuro. No entanto, o evento da imaginação é um evento real, que realmente ocorre, e é isso que tem efeito. É a sua imaginação do evento irreal, sua imaginação muito real, implementada em um cérebro bastante físico, que afeta seu comportamento.

Acreditar que o evento real de um carro atropelando você, um evento que poderia ter acontecido, mas não aconteceu, está exercendo um efeito causal direto em seu comportamento, é confundir o mapa com o território.

O que afeta o mundo é real. (Se as coisas podem afetar o mundo sem serem 'reais', é difícil definir o que a palavra 'real' significa.) As configurações e os fluxos de amplitude são causas e têm efeitos visíveis; eles são reais. As configurações não são mundos possíveis, e as amplitudes não são graus de crença, assim como sua cadeira não é um mundo possível e o céu não é um grau de crença.

Então, o que é uma configuração?

Bem, você terá uma ideia mais clara disso em ensaios posteriores.

Mas para lhe dar uma visão rápida de como a imagem real difere da versão simplificada que vimos neste ensaio…

Nossa configuração experimental lida apenas com uma partícula em movimento, um único fóton. As configurações reais envolvem múltiplas partículas. O próximo ensaio abordará o caso de mais de uma partícula e fornecerá uma compreensão muito mais clara do que são configurações.

Cada configuração sobre a qual falamos deveria descrever uma posição conjunta de todas as partículas nos espelhos e detectores, não apenas a posição de um fóton em movimento.

Na verdade, as configurações realmente reais envolvem as posições conjuntas de todas as partículas no universo, incluindo as partículas que compõem os experimentadores. Você pode ver por que estou guardando a noção de resultados experimentais para ensaios posteriores. No mundo real, a amplitude é uma distribuição contínua em um espaço contínuo de configurações. As 'configurações' deste ensaio eram em blocos e digitais, assim como nossos 'fluxos de amplitude'. Era como se estivéssemos falando de um fóton se teletransportando de um lugar para outro.

Se nada disso fez sentido, não se preocupe. Tudo ficará mais claro nos ensaios posteriores. Queria apenas dar uma ideia de onde isso estava indo.

[1][Nota do editor: estritamente falando, um espelho semi-refletor padrão produziria uma regra “multiplique por -1 quando o fóton gira em um ângulo reto”, não “multiplique por i”. O cenário básico descrito pelo autor não é fisicamente impossível e seu uso não afeta o argumento substantivo. No entanto, os estudantes de física podem ficar confusos se compararem a discussão aqui com as discussões dos livros sobre interferômetros de Mach-Zehnder. Deixamos essa idiossincrasia no texto porque elimina qualquer necessidade de especificar qual lado do espelho é meio prateado, simplificando o experimento.]”

231 - Configurações conjuntas

A chave para compreender as configurações e, consequentemente, a chave para compreender a mecânica quântica reside em perceber, em um nível genuinamente visceral, que as configurações envolvem mais do que uma única partícula.

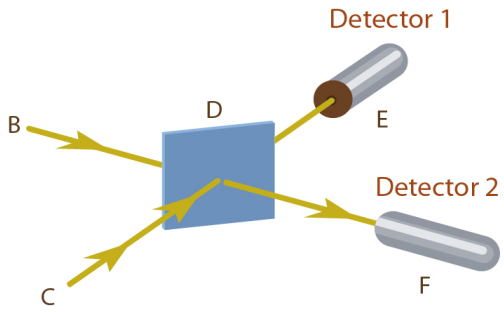


Figura 231.1

Dando continuidade ao ensaio anterior, a Figura 231.1 exibe uma versão modificada do experimento no qual dois fótons são enviados simultaneamente de B e C em direção a D.

A configuração inicial é a seguinte:

“Um fóton indo de B para D,

e um fóton indo de C para D.”

Vamos supor, mais uma vez, que a amplitude da configuração inicial seja (-1 + 0i).

Lembrando a regra do espelho semi-refletor (em D), que indica que uma deflexão em ângulo reto é multiplicada por i e uma trajetória reta é multiplicada por 1.

Então os fluxos de amplitude, a partir da configuração inicial, considerando separadamente os quatro casos de deflexão/não deflexão de cada fóton, são:

1. O fóton de “B para D” é desviado e o fóton de “C para D” também é desviado. Nesse caso, a amplitude flui para a configuração de “um fóton indo de D para E e um fóton indo de D para F”. A amplitude do fluxo é (1 + 0i).
2. O fóton de “B para D” é desviado e o fóton de “C para D” segue em linha reta. Isso leva à amplitude fluindo para a configuração de “dois fótons indo de D para E”. A amplitude do fluxo é (0 - i).
3. O fóton de “B para D” segue em linha reta e o fóton de “C para D” é desviado. Nesse caso, a amplitude flui para a configuração de “dois fótons indo de D para F”. A amplitude do fluxo é (0 - i).
4. O fóton de “B para D” segue reto e o fóton de “C para D” segue reto. Essa amplitude flui para a configuração de “um fóton indo de D para F e um fóton indo de D para E”. A amplitude do fluxo é (-1 + 0i).

Agora, e isso é uma ideia crucial e fundamental na mecânica quântica, as amplitudes nos casos 1 e 4 estão fluindo para a mesma configuração. Se tanto o fóton B quanto o fóton C seguem em linha reta ou ambos são desviados, a configuração resultante é a de um fóton indo em direção a E e outro fóton indo em direção a F.

Portanto, somamos os dois fluxos de amplitude do caso 1 e do caso 4 e obtemos uma amplitude total de (1 + 0i) + (-1 + 0i) = 0.

Quando aplicamos nosso misterioso leitor de módulo quadrado às três configurações finais, descobrimos que "dois fótons no Detector 1" e "dois fótons no Detector 2" têm o mesmo módulo quadrado, enquanto "um fóton no Detector 1 e um fóton no Detector 2" tem um módulo quadrado de zero.

Na experiência prática, nunca encontramos o Detector 1 e o Detector 2 registrando ambos disparos. Encontramos o Detector 1 disparando duas vezes ou o Detector 2 disparando duas vezes, com a mesma frequência. (Supondo que eu tenha entendido matemática e física corretamente, embora eu não tenha realizado o experimento pessoalmente.)

A identidade da configuração não é "o fóton B indo para E e o fóton C indo para F". Nesse caso, as configurações resultantes nos casos 1 e 4 não seriam iguais. No caso 1, teríamos "fóton B para E, fóton C para F", e no caso 4 teríamos "fóton B para F, fóton C para E". Essas seriam duas configurações distintas se as configurações tivessem uma estrutura de rastreamento de fótons.

Portanto, não somaríamos as duas amplitudes e as anularíamos. Manteríamos as amplitudes em duas configurações separadas. As amplitudes totais teriam módulos quadrados diferentes de zero. E quando realizássemos o experimento, descobriríamos (aproximadamente metade das vezes) que o Detector 1 e o Detector 2 registraram um fóton cada. Isso não ocorre se meus cálculos estiverem corretos.

As configurações não rastreiam a origem das partículas. A identidade de uma configuração é simplesmente "um fóton aqui, um fóton ali; um elétron aqui, um elétron ali". Não importa como se chega a essa situação, desde que as mesmas espécies de partículas estejam nos mesmos lugares, isso é considerado a mesma configuração.

Reitero que a pergunta "Que tipo de informação a estrutura da configuração incorpora?" tem implicações experimentais. Podemos deduzir, a partir do experimento, como a própria realidade deve estar tratando as configurações.

Em um universo clássico, não haveria implicações experimentais. Se um fóton fosse como uma pequena bola de bilhar, que fosse para um lado ou outro, e as configurações representassem nossas crenças sobre os possíveis estados do sistema, e em vez de amplitudes tivéssemos probabilidades, não faria diferença se rastreássemos a origem dos fótons ou jogássemos a informação fora.

No universo clássico, eu poderia atribuir uma probabilidade de 25% para cada um dos seguintes cenários:

- Ambos os fótons vão para E.
- Ambos os fótons vão para F.
- O fóton B vai para E, e o fóton C vai para F.
- O fóton B vai para F e o fóton C vai para E.

Como não me importo com qual dos dois últimos casos ocorreu, posso agrupá-los em um só e somar suas probabilidades, declarando assim: "uma probabilidade de 50% de que cada detector receba um fóton".

Com as probabilidades, podemos combinar eventos como quisermos - traçar nossos limites em torno de conjuntos de mundos possíveis como quisermos - e os números ainda funcionarão da mesma maneira. A probabilidade de dois eventos mutuamente exclusivos é sempre igual à probabilidade do primeiro evento mais a probabilidade do segundo evento.

No entanto, você não pode agrupar ou separar configurações arbitrariamente em seu modelo e obter

as mesmas previsões experimentais. Nossa ferramenta mágica nos fornece as razões dos módulos quadrados. Quando somamos dois números complexos, o módulo quadrado da soma não é a soma dos módulos quadrados das partes:

Módulo_Quadrado($C1 + C2$) ≠ Módulo_Quadrado($C1$)+ Módulo_quadrado($C2$)

Por exemplo:

S_M ((2 + i) + (1 − i)) = S_M (3 + 0i)

$= 3^2 + 0^2$

= 9 ,

S_M (2 + i) + S_M (1 − i) = $(2^2 + 1^2) + (1^2 + (-1)^2)$

= (4 + 1) + (1 + 1)

= 7.

No experimento atual, os fluxos de (1 + 0i) e (-1 + 0i) se anularam, resultando em um módulo quadrado de 0, enquanto o módulo quadrado das partes teria sido 1.

Se em vez de Módulo_Quadrado, nossa ferramenta mágica fosse uma função linear - qualquer função na qual F (X + Y) = F (X) + F (Y) - toda a mecânica quântica desapareceria instantaneamente e seria substituída por uma física clássica. (Uma física clássica diferente, não a mesma ilusão de classicidade que imaginamos de dentro dos níveis mais elevados de organização em nosso próprio mundo quântico.)

Se as amplitudes fossem somente probabilidades, elas não poderiam se anular quando os fluxos colidem. Se as configurações fossem apenas estados de conhecimento, você poderia reorganizá-las como quisesse.

Mas as configurações estão cravadas no lugar, são indivisíveis e não podem ser combinadas sem alterar as leis da física.

E parte do que está cravado é como as configurações tratam múltiplas partículas. Uma configuração diz “um fóton aqui, um fóton ali”, não “este fóton aqui, aquele fóton ali”. “Esse fóton aqui, aquele fóton ali” não tem uma identidade diferente de “aquele fóton aqui, esse fóton ali”.

O resultado, como observado no experimento de hoje, é que você não pode desmembrar a física de nosso universo para tratar partículas com identidades individuais.

Uma das razões pelas quais os humanos têm dificuldade em lidar com a física quântica perfeitamente natural é que eles estranhamente continuam tentando decompor a realidade em uma soma de bolas de bilhar individualmente reais.

Ha, ha! Humanos tolos.

232 — Configurações distintas

O experimento no ensaio anterior trouxe duas lições-chave:

Primeiro, vimos que os fluxos de amplitude podem se cancelar, e, porque nossa medida mágica de módulo ao quadrado não é linear, a identidade das configurações é fixada — você não pode reorganizar configurações do mesmo jeito que pode agrupar mundos possíveis. Quais configurações são as mesmas, e quais são distintas, têm consequências experimentais; é um fato observável.

Segundo, vimos que configurações envolvem múltiplas partículas. Se houver dois fótons entrando no aparato, isso não significa que existam duas configurações iniciais. Em vez disso, a identidade da configuração inicial é "dois fótons entrando." (Idealmente, cada configuração que discutimos incluiria todas as partículas no experimento — incluindo as partículas que compõem os espelhos e detectores. E no universo real, cada configuração é sobre todas as partículas… em todos os lugares.

O que torna as configurações distintas não são partículas distintas. Cada configuração envolve todas as partículas. O que torna as configurações distintas é que as partículas ocuparem posições diferentes — pelo menos uma partícula em um estado diferente.

Para dar uma demonstração importante…

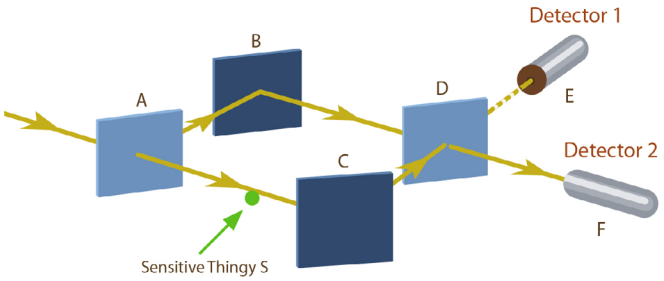


Figura 232.1

A Figura 232.1 é o mesmo experimento da Figura 230.2, com uma mudança importante: entre A e C foi colocado um dispositivo sensível, S. A principal característica de S é que, se um fóton passar por S, então S termina em um estado ligeiramente diferente.

Digamos que os dois possíveis estados de S são Sim e Não. O dispositivo sensível S começa no estado Não, e termina no estado "Sim" se um fóton passar.

Então, a configuração inicial é: "fóton indo para A; e S no estado Não," (−1 + 0i).

Em seguida, a ação do espelho semi-refletor em A. Na versão anterior deste experimento, sem o dispositivo sensível, as duas configurações resultantes foram "A para B" com amplitude −i e "A para C" com amplitude −1. Agora, no entanto, um novo elemento foi introduzido no sistema, e todas as configurações são sobre todas as partículas, então cada configuração menciona o novo elemento. Assim, os fluxos de amplitude

da configuração inicial são para:

"fóton de A para B; e S no estado Não," (0 − i)

"fóton de A para C; e S no estado Sim," (−1 + 0i):

Em seguida, a ação dos espelhos completos em B e C:

"fóton de B para D; e S no estado Não," (1 + 0i)

"fóton de C para D; e S no estado Sim," (0 − i):

E então a ação do espelho semi-refletor em D, sobre a amplitude fluindo de ambas as configurações acima:

1. "fóton de D para E; e S no estado Não," (0 + i)
2. "fóton de D para F; e S no estado Não," (1 + 0i)
3. "fóton de D para E; e S no estado Sim," (0 − i)
4. "fóton de D para F; e S no estado Sim," (1 + 0i):

Quando fizemos este experimento sem o dispositivo sensível, os fluxos de amplitude (1) e (3) de (0 + i) e (0 − i) para a configuração "D para E" se cancelaram mutuamente. Ficamos sem amplitude para um fóton indo para o Detector 1 (ao nível experimental, nunca observamos um fóton atingindo o Detector 1).

Mas neste caso, os dois fluxos de amplitude (1) e (3) agora são para configurações distintas; pelo menos uma entidade, S, está em um estado diferente entre (1) e (3). As amplitudes não se cancelam.

Quando passamos nosso detector mágico de módulo ao quadrado sobre as quatro configurações finais, descobrimos que os módulos quadrados de todas são iguais: 25% de probabilidade cada. Ao nível de mundo real, descobrimos que o fóton tem chances iguais de atingir o Detector 1 e o Detector 2.

Tudo isso é verdade, mesmo que nós, os pesquisadores, não nos importemos com o estado de S. Ao contrário dos mundos possíveis, as configurações não podem ser reagrupadas ao acaso. As leis da física dizem que as duas configurações são distintas; não é uma questão de como podemos mais convenientemente dividir o mundo.

Tudo isso é verdade, mesmo que não nos preocupemos em olhar o estado de S. As configurações (1) e (3) são distintas na física, mesmo que não conheçamos a distinção.

Tudo isso é verdade, mesmo que não saibamos que S existe. As configurações (1) e (3) são distintas, quer tenhamos ou não representações mentais distintas para as duas possibilidades.

Tudo isso é verdade, mesmo que estejamos no espaço, e S transmita um novo fóton em direção ao vazio interestelar em duas direções distintas, dependendo de o fóton de interesse ter passado por ele ou não. De modo que nunca poderíamos descobrir se S estava em "Sim" ou "Não". O estado de S seria incorporado no fóton transmitido para lugar nenhum. O fóton perdido pode ser um invisível implícito, e o estado de S pragmaticamente indetectável; mas as configurações ainda são distintas.

(A principal razão pela qual isso não funcionaria, é se S fosse empurrado, mas S tinha uma distribuição original no espaço de configurações que era maior do que o empurrão. Então você não poderia confiar no empurrão para separar a distribuição de amplitude sobre o espaço de configurações em aglomerados distintos. Na realidade, tudo isso acontece numa distribuição de amplitude diferenciável sobre um espaço de configurações contínuo.)

Configurações não são estados de crença. Sua distinção é um fato objetivo com consequências experimentais. As configurações são distintas mesmo se ninguém souber o estado de S; distintas mesmo se nenhuma entidade inteligente puder descobrir. As configurações são distintas desde que pelo menos uma partícula no universo esteja em uma posição diferente. Isso é demonstrável experimentalmente.

Por que estou enfatizando isso? Porque, na era das trevas, quando ninguém entendia física quântica…

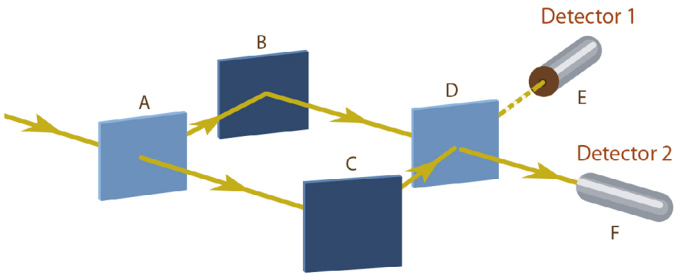


Figura 232.2

Ok, então imagine que você não tem a menor ideia do que realmente está acontecendo, e você tenta o experimento na Figura 232.2, e nenhum fóton aparece no Detector 1. Legal.

Você também descobre que quando coloca um bloqueio entre B e D, ou um bloqueio entre A e C, os fótons aparecem nos Detectores 1 e 2 em proporções iguais. Mas apenas um de cada vez — ou o Detector 1, ou o Detector 2 dispara, não ambos simultaneamente.

Então, sim, parece que você está lidando com uma partícula — o fóton está apenas em um lugar de cada vez, toda vez que você o vê.

E ainda assim, há algum tipo de… fenômeno misterioso… que impede o fóton de aparecer no Detector 1. E esse fenômeno misterioso depende de o fóton conseguir ir por ambos os caminhos. Mesmo que o fóton só apareça em um detector ou no outro, mostrando, você pensaria que o fóton está apenas em um lugar de cada vez.

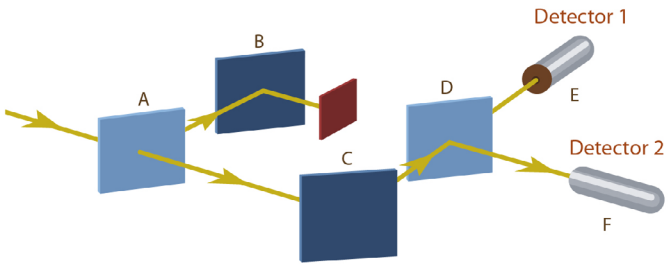


Figura 232.3

O que torna o padrão dos experimentos bem bizarro! Afinal, o fóton ou vai de A para C, ou de A para B; um ou o outro. (Ou assim você pensaria, se estivesse instintivamente tentando dividir a realidade em partículas individualmente reais.) Mas quando você bloqueia um curso ou outro, como na Figura 232.3, você começa a obter resultados experimentais diferentes!

É como se o fóton quisesse ser permitido ir por ambos os caminhos, mesmo que (você pensaria) ele vá por um caminho ou pelo outro. E ele pode perceber se você tentar bloqueá-lo, sem realmente ir lá — se ele tivesse ido, teria encontrado o bloqueio e não atingido nenhum detector.

É como se meras possibilidades pudessem ter efeitos causais, em desafio ao que a palavra “real” geralmente significa…

Mas é um pouco cedo para pular para conclusões como essa, quando você não tem uma imagem completa do que acontece no experimento.

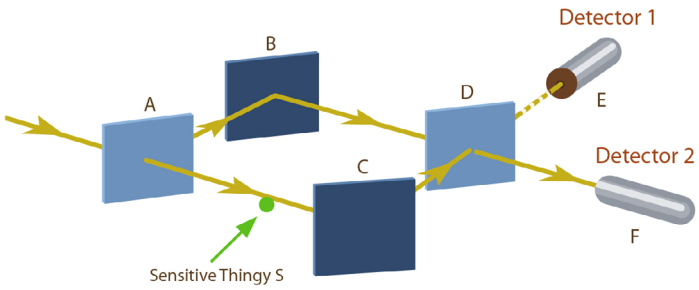


Figura 232.4

Então, ocorre a você colocar um sensor entre A e C, como na Figura 232.4, para que você possa saber qual caminho o fóton realmente toma em cada ocasião.

E o fenômeno misterioso desaparece.

Quero dizer, agora, quão louco é isso? Que tipo de paranoia isso inspira em algum pobre cientista?

Ok, então no século XXI percebemos que, para "saber" a história de um fóton, as partículas que compõem seu cérebro precisam estar correlacionadas com a história do fóton. Se um pequeno dispositivo sensível, S, que está correlacionado com a história do fóton, for suficiente para diferenciar as configurações finais e impedir que os fluxos de amplitude se cancelem, então um sensor completo com um visor digital, ou mesmo um cérebro humano, colocaria septilhões de partículas em diferentes posições e impediria que os fluxos de amplitude de interferência se cancelassem.

Mas se você ainda não tivesse descoberto isso…

Então, você ponderaria sobre o sensor que baniu o Fenômeno Misterioso, e pensaria:

O fóton não quer apenas estar fisicamente livre para ir por qualquer caminho. Não é uma pequena onda seguindo um caminho desimpedido, porque então apenas ter um caminho fisicamente desimpedido seria suficiente.

Não… Eu não posso saber por qual caminho o fóton foi.

O fenômeno misterioso… não quer que eu olhe muito de perto… enquanto ele faz sua coisa misteriosa.

Não são as possibilidades físicas que têm um efeito na realidade… apenas as possibilidades epistêmicas. Se sei por qual caminho o fóton foi, já não é mais plausível que ele tenha ido pelo outro caminho… interrompendo o fenômeno misterioso tão efetivamente quanto colocar um bloqueio entre B e D.

Eu tenho que não observar qual caminho o fóton foi, para garantir que ele sempre chegue ao Detector 2. Deve ser razoável que o fóton pode ter seguido qualquer uma das trajetórias possíveis, seja para B ou para C. O que posso saber é o fator determinante, independentemente de quais caminhos físicos eu escolho abrir ou fechar.

PAREM AS MÁQUINAS! A MENTE É FUNDAMENTAL AFINAL! A CONSCIÊNCIA DETERMINA NOSSOS RESULTADOS EXPERIMENTAIS!

Você ainda pode ler esse tipo de coisa. Em livros didáticos de física. Mesmo agora, quando uma maioria dos físicos teóricos sabe melhor. Pare as máquinas. Por favor, pare as máquinas.

O retrospecto é 20/20; então é fácil dizer que, em retrospectiva, havia certas pistas de que essa interpretação não estava correta.

Por exemplo, se você colocar o sensor entre A e C, mas não o ler, o fenômeno misterioso ainda desaparece, e o fóton ainda às vezes termina no Detector 1. (Ah, mas você poderia ter lido, e agora as possibilidades são reais…)

Mas nem precisa ser um sensor, um instrumento científico que você construiu. Uma única partícula empurrada o suficiente dissipará a interferência. Um fóton irradiando para onde você nunca mais o verá pode fazer o truque. Não há muita participação humana aí. Não há muita consciência envolvida.

Talvez antes de soar o alarme dualista de que os cérebros humanos são fisicamente especiais, você deva fornecer uma prova experimental de que uma pedra não pode desempenhar o mesmo papel em dissipar o Fenômeno Misterioso como um pesquisador humano?

Mas isso é retrospectiva, e é fácil fazer julgamentos com retrospectivas. Você realmente acha que poderia ter feito melhor do que John von Neumann, se estivesse vivo na época? O objetivo desse tipo de análise retrospectiva é perguntar quais tipos de pistas totalmente gerais você poderia ter seguido, e se há pistas semelhantes que você está ignorando agora em mistérios atuais.

Embora seja um pouco embaraçoso que, mesmo após a teoria das amplitudes e configurações ter sido desenvolvida — com a teoria agora dando a previsão definitiva de que qualquer partícula empurrada faria o truque —, os primeiros cientistas ainda não entenderam.

Mas veja… havia sido estabelecido como Sabedoria Comum que as configurações eram possibilidades, era a possibilidade epistêmica que importava, amplitudes eram um tipo muito estranho de informação parcial, e a observação consciente fazia a quanticidade desaparecer. E que era melhor evitar pensar muito sobre todo esse assunto, desde que suas previsões experimentais dessem certo.

233 - Postulados de colapso

A decoerência macroscópica, também conhecida como “muitos mundos”, é a ideia de que as leis quânticas conhecidas que regem eventos microscópicos, simplesmente regem em todos os níveis sem alterações. Na época em que as pessoas não conheciam a decoerência - antes que ocorresse a alguém que as leis deduzidas com tanta precisão para a física microscópica pudessem se aplicar universalmente - o que as pessoas acreditavam que estava acontecendo?

O raciocínio inicial era mais ou menos assim:

Quando meus cálculos mostraram uma amplitude de (−1/3)i para a absorção deste fóton, minhas estatísticas experimentais mostraram que o fóton foi absorvido cerca de 107 vezes em 1.000, o que se ajusta bem a 1/9, o quadrado do módulo.

Isso levou a:

A amplitude é a probabilidade (através do módulo quadrado).

E depois a:

Quando você mede algo e sabe que não aconteceu, sua probabilidade se torna zero.

Lido literalmente, isso implica que o conhecimento - ou até mesmo a percepção consciente - causa o colapso. Essa foi, de fato, a forma da teoria apresentada por Werner Heisenberg!

Mas as pessoas ficaram cada vez mais inquietas com a noção de importar ideias dualistas na física fundamental. E com razão! Então, substituíram o argumento original por um colapso “objetivo”, que destruía todas as partes da função de onda, exceto uma, e era acionado em algum momento antes da superposição atingir níveis na escala humana.

Ao supor que partes da função de onda podem sumir, você poderia perguntar:

Existe apenas um sobrevivente? Talvez muitos mundos sobrevivam. Sua frequência poderia corresponder ao seu módulo quadrado integrado. Assim, o mundo sobrevivente típico seguiria a regra de Born.

Mas as teorias de colapso modernas só sugerem um mundo sobrevivente. Por quê?

As teorias do colapso foram criadas em uma época em que simplesmente não ocorria a nenhum físico a possibilidade de existir mais de um mundo! As pessoas tinham como certo que as medições tinham resultados únicos - era uma suposição tão profunda, que era invisível, porque era o que elas viam acontecendo. As teorias do colapso foram criadas para explicar por que as medições tinham resultados únicos, em vez de (em geral) explicar por que as estatísticas experimentais correspondiam à regra de Born.

Por motivos semelhantes, os “postulados de colapso” considerados academicamente supõe que o colapso acontece antes dos humanos ficarem superpostos. Mas os experimentos continuam descartando o “colapso” em sistemas emaranhados crescentes. Aparentemente, há um experimento planejado para demostrar a superposição quântica em escalas de 50 micrômetros. Isso é maior que a maioria dos neurônios e quase do tamanho de alguns fios de cabelo!

Então, por que alguém não se adianta e pergunta:

Digamos que continuamos tendo que postular que o colapso acontece cada vez mais tarde. E se o colapso só ocorrer quando a superposição atingir escalas planetárias e ocorrer uma divergência substancial - digamos que a função de onda da Terra colapsa cerca de uma vez por minuto? Então, as Terras sobreviventes em um determinado momento lembrariam de uma longa história de experimentos quânticos seguindo as estatísticas de Born, uma grande maioria começaria a obter resultados não-Born de experimentos quânticos e, em seguida, deixaria de existir abruptamente um minuto depois.

Por que teorias de colapso assim não têm muitos seguidores acadêmicos, entre as muitas pessoas que aparentemente acham que não há problema em partes da função de onda simplesmente desaparecerem? ? Especialmente considerando que os experimentos estão provando a superposição em sistemas cada vez maiores?

Um cínico poderia sugerir que o motivo do apoio contínuo ao colapso não é a plausabilidade física de ter grandes partes da função de onda desaparecendo repentinamente, ou a esperança de explicar de alguma forma as estatísticas de Born. O objetivo é manter o apelo intuitivo de que "não me lembro da medição ter mais de um resultado, então só uma coisa aconteceu; não me lembro de me dividir, então só deve existir um de mim". Você não se lembra de morrer, então humanos superpostos nunca devem colapsar. Uma teoria que ignorasse a intuição perderia o ponto principal. Você poderia muito bem passar para a decoerência.

É o que um cínico poderia sugerir.

Mas com certeza é cedo demais para atacar os motivos dos apoiadores do colapso. Isso é apenas ad hominem. E quanto à plausabilidade física real das teorias de colapso?

Primeiro: Alguma teoria do colapso tem apoio experimental? Não.

Esclarecido isso…

Se o colapso funcionasse como seus defensores afirmam, seria:

1. A única evolução não-linear em toda a mecânica quântica.
2. A única evolução não-unitária em toda a mecânica quântica.
3. O único fenômeno não-diferenciável (na verdade, descontínuo) em toda a mecânica quântica.
4. O único fenômeno em toda a mecânica quântica não-local no espaço de configuração.
5. O único fenômeno em toda a física que viola a simetria CPT.
6. O único fenômeno em toda a física que viola o Teorema de Liouville (tem um mapeamento de muitos para um das condições iniciais para os resultados).
7. O único fenômeno em toda a física que é não causal / não-determinístico / inerentemente aleatório.
8. O único fenômeno em toda a física que é não-local no espaço-tempo e propaga influência mais rápido que a luz.

O QUE O MALDITO POSTULADO DO COLAPSO PRECISA FAZER PARA OS FÍSICOS O REJEITAREM? MATAR UM MALDITO FILHOTE?

234 - A decoerência é simples

Uma carta aos físicos:

Quando eu era apenas um garotinho, meu pai, um físico com doutorado, advertiu-me severamente contra a intromissão nos assuntos dos físicos; ele disse ser inútil tentar compreender a física sem a matemática formal. Ponto final. Sem cláusulas de escape. Mas li nos livros populares de Feynman que, se você entendesse realmente de física, deveria conseguir explicá-la a um não físico. Acreditei em Feynman em vez de em meu pai, porque Feynman ganhou o Prêmio Nobel e meu pai não.

Foi só mais tarde - quando estava lendo as Palestras de Feynman, na verdade - que percebi que meu pai havia me dado a verdade simples e honesta. Sem matemática = sem física.

Por vocação, sou um bayesiano, não um físico. No entanto, embora tenha sido criado para não me intrometer nos assuntos dos físicos, minha mão foi forçada pelo ocasional mau uso grosseiro de três termos: simples, falsificável e testável.

A introdução anterior é para que você não ria e diga: "Claro que sei o que essas palavras significam!" Há matemática aqui. O que se segue será uma reafirmação dos pontos em Crença no Invisível Implícito, conforme eles se aplicam à física quântica.

Comecemos com a observação que me iniciou por essa jornada, da qual vi várias versões; parafraseando, ela é assim:

A interpretação de muitos mundos da mecânica quântica postula que há um grande número de outros mundos, coexistindo com o nosso. A Navalha de Occam diz que não devemos multiplicar entidades desnecessariamente.

Agora, deve ser dito, com toda justiça, que aqueles que dizem isso geralmente também confessam:

Mas esta não é uma aplicação universalmente aceita da Navalha de Ocam; alguns dizem que a Navalha de Occam deve ser aplicada às leis que regem o modelo, não ao número de objetos no modelo.

Portanto, é bom que todos reconheçamos os argumentos contrários e consideremos ambos os lados da história.

Mas suponha que você precise calcular a simplicidade de uma teoria.

A formulação original de Guilherme de Ockham afirmava:

Lex parsimoniae: Entia non sunt multiplicanda praeter necessitatem.

"Lei da Parcimônia: Entidades não devem ser multiplicadas além do necessário".

No entanto, isso é um conselho qualitativo. Não basta dizer se uma teoria parece mais simples ou mais complexa do que outra - você precisa atribuir um número; e o número precisa ter significado, não pode ser inventado. Superar essa lacuna é como a diferença entre poder observar quais coisas estão se movendo "rápido" ou "devagar" e começar a medir e calcular velocidades.

Suponha que você tenha tentado dizer: “Conte as palavras - é assim que se mede a complexidade de uma teoria”.

Robert Heinlein certa vez afirmou (com ironia, espero) que a “explicação mais simples” é sempre: “A mulher na rua é uma bruxa; foi ela que fez.” Onze palavras - não há muitos artigos de física que superem isso.

Diante desse desafio, existem dois caminhos diferentes que você pode seguir.

Primeiro, você pode perguntar: “A mulher lá da rua é o quê?” Apenas porque o inglês possui uma palavra para indicar um conceito, não significa que o próprio conceito seja simples. Suponha que você estivesse conversando com alienígenas que não conhecem bruxas, mulheres ou ruas, quanto tempo levaria para explicar sua teoria a eles? Melhor ainda, suponha que você tivesse que escrever um programa que incorporasse sua hipótese e gerasse o que você afirma serem as previsões de sua hipótese - quão grande esse programa teria que ser? Digamos que sua tarefa seja prever uma série temporal de posições medidas para uma pedra rolando morro abaixo. Se você escrever uma sub-rotina que simula bruxas, isso não parece ajudar a reduzir o local para onde a pedra rola - a sub-rotina adicional apenas infla seu código. No entanto, você pode descobrir que seu código inclui necessariamente uma sub-rotina que calcula o quadrado de números.

Em segundo lugar, você pode perguntar: “A mulher lá da rua é uma bruxa; ela fez o quê? Suponha que você queira descrever um evento com máxima precisão, considerando as evidências disponíveis - digamos, o movimento de uma pedra rolando morro abaixo ao longo do tempo. Você pode iniciar sua explicação afirmando: ‘Aquela mulher na rua é uma bruxa.’ Entretanto, seu amigo logo indaga: ‘O que ela fez?’ Você, então, prossegue com: ‘Ela fez a pedra rolar um metro após o primeiro segundo, nove metros após o terceiro segundo…’ Inserir ‘A mulher na rua é uma bruxa’ no início da mensagem não facilita a compressão do restante da descrição. De modo geral, acaba resultando em uma mensagem mais extensa do que o necessário. Portanto, faz mais sentido omitir a referência à ‘bruxa’. Por outro lado, se você introduzir a história de Galileu, poderá reduzir significativamente a quantidade de informações necessárias para descrever as próximas cinco mil séries temporais detalhadas relacionadas a pedras rolando morro abaixo.

Se você optar pela primeira abordagem, isso nos leva ao conceito de complexidade de Kolmogorov e indução de Solomonoff. Seguir o segundo caminho nos leva ao conceito de Comprimento Mínimo da Mensagem.

Ah, então eu posso escolher entre essas definições de simplicidade?

Na verdade, não; em suas formas mais avançadas, ambos os formalismos se mostraram equivalentes.

E, suponho que agora você argumentará que ambos os formalismos se alinham com o princípio de ‘Occam conta leis, não objetos’.

Mais ou menos. No Comprimento Mínimo da Mensagem, desde que você forneça ao seu amigo uma receita exata, ele poderá segui-la mentalmente para obter a série temporal do movimento da pedra, não nos importamos com a quantidade de esforço mental necessária para seguir a receita. Já na indução de Solomonoff, contamos os bits no código do programa, não os bits de RAM usados enquanto o programa é executado. Nesse contexto, ‘entidades’ são linhas de código, não objetos físicos. E, como mencionado anteriormente, esses dois formalismos, em última análise, são equivalentes.

Agora, antes de entrar em mais detalhes sobre a simplicidade formal, permita-me fazer uma digressão para considerar a objeção:

‘E daí? Por que eu não posso simplesmente inventar meu próprio formalismo que aborde as coisas de maneira diferente? Por que devo adotar a abordagem que você escolheu para o seu campo? Existe alguma evidência experimental que respalda essa abordagem?’

Na verdade, sim, acredite ou não. No entanto, antes de entrar nesse assunto, proponho começarmos do início.

A regra da conjunção na teoria da probabilidade afirma:

P (X, Y) ≤ P (X).

Para quaisquer proposições X e Y, a probabilidade de que 'X seja verdadeiro, e Y seja verdadeiro' é menor ou igual à probabilidade de que 'X, seja verdadeiro (independentemente de Y ser verdadeiro ou falso)'. (Se isso não parecer muito profundo à primeira vista, saiba que existem situações em que os seres humanos que avaliam probabilidades violam essa regra).

Normalmente, a regra de conjunção P (X, Y) ≤ P (X) não pode ser aplicada diretamente a um conflito entre hipóteses mutuamente exclusivas. Ela se aplica apenas nos casos em que o lado esquerdo implica estritamente o lado direito. Além disso, a conjunção é apenas uma desigualdade; não fornece o tipo de cálculo quantitativo que desejamos.

Entretanto, a regra de conjunção nos dá uma regra de diminuição monotônica na probabilidade: à medida que você acrescenta detalhes a uma narrativa, e cada detalhe adicional pode ser verdadeiro ou falso, a probabilidade da narrativa diminui monotonicamente. Pense na probabilidade como uma quantidade conservada: não há muito o que fazer. À medida que o número de detalhes em uma narrativa aumenta, o número de narrativas possíveis aumenta exponencialmente, mas a soma de suas probabilidades nunca pode ultrapassar 1. Para cada narrativa 'X e Y', há uma narrativa 'X e ¬Y.' Quando você simplesmente relata 'X', você pode somar as probabilidades de Y e ¬Y.

Se você acrescentar dez detalhes a X, cada um dos quais pode ser potencialmente verdadeiro ou falso, então essa narrativa deve competir com outras 2^{10} - 1 narrativas igualmente detalhadas por uma probabilidade limitada. Por outro lado, ao mencionar apenas 'X', você pode somar sua probabilidade em 2^{10} narrativas

((X e Y e Z e...) ou (X e ¬Y e Z e...) ou ...).

As 'entidades' consideradas pela Navalha de Occam devem ser individualmente caras em probabilidade; é por isso que preferimos teorias com menos delas.

Imagine uma loteria que vende um milhão de bilhetes, onde cada bilhete é vendido apenas uma vez, e todos os bilhetes foram vendidos antes do sorteio. Seu amigo comprou um bilhete por US$ 1 - um investimento que parece ruim, pois o prêmio é de apenas US$ 500.000. No entanto, seu amigo argumenta: 'Considere as hipóteses alternativas, 'Amanhã, alguém ganhará na loteria' e 'Amanhã, eu ganharei na loteria'. Claramente, a última hipótese é mais simples, conforme a Navalha de Ocam; ela menciona apenas uma pessoa e um bilhete, enquanto a primeira hipótese é mais complexa, falando de um milhão de pessoas e um milhão de bilhetes!'

Dizer que a Navalha de Ocam conta apenas as leis e não os objetos, não é inteiramente preciso: o que é contado contra uma teoria são as entidades que ela deve mencionar explicitamente, uma vez que essas entidades não podem ser agregadas. Suponha que você e seu amigo estejam intrigados com uma jogada de bilhar extraordinária, na qual você é informado sobre o estado inicial da mesa de bilhar e quais bolas foram encaçapadas, mas não sabe como a jogada foi realizada. Você apresenta uma teoria que inclui dez colisões específicas entre dez bolas específicas; seu amigo propõe uma teoria que envolve cinco colisões específicas entre cinco bolas específicas. O que conta contra suas teorias não são apenas as leis que afirmam governar as bolas de bilhar, mas também as bolas de bilhar específicas que precisariam estar em estados específicos para que as previsões de seus modelos fossem bem-sucedidas.

Se você medir a temperatura de uma sala como sendo 22 °C, não faz sentido argumentar: 'Seu termômetro está provavelmente errado; é muito mais provável que a temperatura seja de 20 °C. Porque, ao considerar todas as partículas na sala, há exponencialmente mais estados que podem ser ocupados se a temperatura for de fato de 22 °C, tornando qualquer estado específico ainda mais improvável.' Mas não importa qual seja o estado exato da sala a 22 °C; você pode fazer a mesma previsão (para a grande maioria desses estados) de que o termômetro continuará marcando 22 °C, e, portanto, você não é sensível ao valor exato das condições iniciais. Você não precisa especificar a posição exata de todas as moléculas de ar na sala, pois isso não afeta a probabilidade da sua explicação.

Por outro lado, voltando ao exemplo da loteria, suponha que seu amigo ganhe dez loterias consecu-

tivas. Nesse ponto, você deveria suspeitar que a questão esteja resolvida. A hipótese de 'Meu amigo sempre ganha na loteria' é mais complexa do que a hipótese de 'Alguém sempre ganha na loteria'. No entanto, a primeira hipótese faz previsões muito mais precisas dos dados, considerando sua longa sequência de vitórias.

No formalismo do Comprimento Mínimo da Mensagem, afirmar 'Há uma única pessoa que ganha na loteria todas as vezes' no início da mensagem comprime a descrição das próximas dez loterias. Você só precisa dizer 'E essa pessoa é Fred Smith' para concluir a mensagem. Compare com a alternativa de especificar cada vitória individualmente: 'A primeira loteria foi ganha por Fred Smith, a segunda loteria foi ganha por Fred Smith, a terceira loteria foi…'

No formalismo da indução de Solomonoff, a probabilidade a priori de 'Meu amigo sempre ganha na loteria' é baixa, porque o programa que descreve a loteria agora requer um código explícito que identifica seu amigo. No entanto, esse programa pode gerar uma distribuição de probabilidade mais estreita sobre os possíveis vencedores da loteria do que 'Alguém sempre ganha na loteria'. Portanto, pela Regra de Bayes, ele pode superar sua baixa probabilidade anterior e emergir como a hipótese mais provável.

Qualquer teoria formal da Navalha de Occam deve definir quantitativamente não apenas 'entidades' e 'simplicidade', mas também o conceito de 'necessidade'.

O Comprimento Mínimo da Mensagem define a necessidade como 'aquilo que comprime a mensagem'.

A indução de Solomonoff atribui uma probabilidade a priori a cada programa possível, com a distribuição geral, sobre cada programa possível, somando no máximo 1. Isso é realizado usando um código binário no qual nenhum software válido é um prefixo de outro programa válido (um 'código sem prefixo'), graças à inclusão de um código de parada. Portanto, a probabilidade a priori de qualquer programa P é simplesmente $2^{(-L(P))}$, em que L(P) representa o comprimento de P em bits.

O próprio programa P pode ser um programa que recebe uma sequência de bits (possivelmente de comprimento zero) e gera a probabilidade condicional de que o próximo bit seja 1, transformando P em uma distribuição de probabilidade sobre todas as sequências binárias. Esse aspecto da indução de Solomonoff, aplicado a qualquer sequência, produz uma mistura de probabilidades posteriores dominada pelos programas mais curtos que preveem a sequência com maior precisão. A soma dessa mistura fornece uma previsão para o próximo bit.

O resultado é que hipóteses mais complexas exigem mais evidências bayesianas - previsões mais precisas ou previsões bem-sucedidas - para justificar hipóteses mais complexas. Mas pode ser feito; o fardo da probabilidade a priori não é infinito. Se você lançar uma moeda quatro vezes e todas as vezes der cara, não concluirá imediatamente que a moeda sempre produzirá cara. Mas se a moeda der cara vinte vezes seguidas, você deve levar a hipótese muito a sério. E quanto à hipótese de que a moeda está programada para produzir uma sequência cíclica de Cara-Coroa-Coroa-Cara-Coroa-Coroa…? Essa é uma ideia mais estranha, mas após cem jogadas consecutivas, seria irracional negá-la.

A química padrão afirma que em um grama de gás hidrogênio existem seiscentos bilhões de trilhões de átomos de hidrogênio. Isso é uma afirmação surpreendente, mas havia evidências suficientes para convencer os físicos em geral, incluindo você, de que isso é verdade.

Agora, se pergunte quanto evidência seria necessária para convencê-lo de uma teoria com seiscentos bilhões de trilhões de leis físicas distintas.

Por que a probabilidade a priori de um programa, no formalismo de Solomonoff, não inclui a quantidade de RAM que o programa utiliza ou o tempo total de execução?

A resposta simples é: 'Porque os recursos de espaço e tempo usados por um programa não são possibilidades mutuamente exclusivas'. Isso difere da especificação do programa, que só pode conter 1 ou 0 em locais específicos.

A resposta ainda mais simples é: 'Porque, historicamente falando, essa heurística não se mostrou

eficaz'.

A Navalha de Ocam surgiu originalmente como uma objeção à sugestão de que as nebulosas eram, na verdade, galáxias distantes - parecia que isso multiplicava enormemente o número de entidades no universo. Todas aquelas estrelas!

Repetidamente, no decorrer da história humana, o universo se revelou cada vez mais vasto. Variantes da Navalha de Ocam que, em cada uma dessas ocasiões, teriam rotulado o universo maior como mais improvável não se saíram bem na experiência histórica da humanidade.

Isso também faz parte da 'evidência experimental' à qual me referi anteriormente. Embora seja possível justificar teorias da simplicidade com base em fundamentos matemáticos, é igualmente importante que funcionem efetivamente, na prática. (A outra parte da 'evidência experimental' vem de estatísticos, cientistas da computação e pesquisadores de inteligência artificial que testam quais definições de 'simplicidade' permitem construir modelos de computador que tenham bom desempenho na previsão de dados futuros com base em dados passados. Provavelmente, o paradigma do Comprimento Mínimo da Mensagem se mostrou mais produtivo nesse sentido, ao ser altamente adaptável na abordagem de problemas reais.)

Imagine uma nave espacial cujo lançamento você testemunha com grande entusiasmo; ela acelera se afastando de você e logo está viajando a 90% da velocidade da luz. Segundo a cosmologia atual, se a expansão do universo continuar, chegará um ponto no futuro em que - de acordo com seu modelo de realidade - você não espera poder interagir com a espaçonave, nem mesmo em princípio. A espaçonave ultrapassou o horizonte cosmológico em relação a você, e os fótons emitidos por ela não conseguirão vencer a expansão do universo.

Você acredita que a espaçonave literalmente desaparece do universo assim que cruza o horizonte cosmológico em relação a você?

Se você acredita que a Navalha de Ocam conta os objetos em um modelo, então, sim, você deveria acreditar nisso. Assim que a espaçonave ultrapassa seu horizonte cosmológico, os modelos em que a espaçonave desaparece instantaneamente e os modelos em que ela continua a avançar fornecem previsões indistinguíveis, e nenhum tem vantagem bayesiana evidencial sobre o outro. No entanto, um modelo contém muitas menos 'entidades', pois não é necessário descrever todos os quarks, elétrons e campos que compõem a espaçonave. Portanto, é mais simples supor que a espaçonave desapareça.

Alternativamente, você poderia argumentar: 'Com base em experimentos anteriores, generalizei certas leis que governam as partículas observadas. A espaçonave é composta por essas partículas. Aplicando essas leis, deduzo que a espaçonave deve continuar após cruzar o horizonte cosmológico, com o mesmo momento e energia que tinha antes. Caso contrário, isso violaria as leis de conservação que observei funcionando em todas as instâncias examinadas. Para supor que a espaçonave desapareça, eu teria que acrescentar uma nova lei: 'As coisas desaparecem assim que cruzam meu horizonte cosmológico'.'

A versão da decoerência (também conhecida como 'muitos mundos') da mecânica quântica afirma que as medidas seguem as mesmas regras da mecânica quântica que governam todos os outros processos físicos. Essa abordagem aplica essas regras a objetos macroscópicos da mesma maneira que a objetos microscópicos, resultando em observadores em estados de superposição. Agora surgem muitas perguntas, como

'Por que nem todas as medidas quânticas binárias parecem ter uma probabilidade de 50/50, já que diferentes versões de nós, veem os dois resultados?'

Entretanto, a objeção de que a decoerência viola a Navalha de Ocam devido à multiplicação de objetos no modelo está simplesmente equivocada

A decoerência não exige que a função de onda comece em um estado complicado específico. Muitos mundos não envolvem especificar manualmente cada mundo, mas os geram com base nas leis compactas da mecânica quântica. Um programa que simule diretamente a mecânica quântica para fazer previsões experimentais, exigiria uma grande quantidade de memória para ser executado, mas simular a função de onda é exponencialmente caro em qualquer abordagem da mecânica quântica! A decoerência simplesmente produz

mais. Muitas descobertas na história, desde as estrelas e galáxias até os átomos e a mecânica quântica, aumentaram significativamente a carga aparente da CPU do que consideramos ser o universo.

Muitos mundos não se traduzem em 'um zilhão de mundos complicados', da mesma forma que a hipótese atômica não se traduz em 'um zilhão de átomos complicados'. Para qualquer pessoa com uma compreensão quantitativa da Navalha de Occam, essa simplesmente não é a interpretação do termo 'complicado'.

Assim como no caso histórico das galáxias, pode ser que as pessoas tenham confundido sua surpresa com a ideia de um universo tão vasto como uma penalidade de probabilidade e tenham invocado a Navalha de Occam para justificá-la. No entanto, se existirem penalidades de probabilidade para a decoerência, a imensidão do universo, em si, definitivamente não é sua fonte!

A noção de que os mundos decoerentes são entidades adicionais penalizadas pela Navalha de Occam está simplesmente equivocada. Não é mais ou menos correta. Não se trata de um argumento fraco, mas ainda válido. Não é uma posição defensável que possa ser sustentada com outros argumentos. É completamente inadequada como teoria da probabilidade. Não é passível de correção. É matemática ruim. 2 + 2 = 3.

235 - A decoerência é falsificável e testável

As palavras “falsificável” e “testável” são às vezes usadas de forma intercambiável, cuja imprecisão é o preço de se comunicar em inglês. Existem duas qualidades teóricas de probabilidade diferentes que desejo discutir aqui, e vou me referir a uma como “falsificável” e a outra como “testável” porque parece o melhor ajuste.

Quanto à matemática, ela começa, como muitas coisas, com:

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}$$

Este é o Teorema de Bayes. Possuo pelo menos duas peças de roupa distintas impressas com este teorema, então deve ser importante.

Para revisar rapidamente, B aqui se refere a um item de evidência, A_i é alguma hipótese sob consideração, e A_j são hipóteses concorrentes e mutuamente exclusivas. A expressão $P(B|A_i)$ significa “a probabilidade de ver B, se a hipótese A_i for verdadeira,” e $P(A_i|B)$ significa “a probabilidade de hipótese A_i ser verdadeira, se virmos B.”

O fenômeno matemático que chamarei de “falsificabilidade” é a propriedade cientificamente desejável de uma hipótese que deve concentrar sua probabilidade em resultados preferidos, implicando que também deve atribuir baixa probabilidade a alguns resultados não desejados. As probabilidades devem somar 1, e existe simplesmente tanta probabilidade para investigar. Idealmente, deve haver observações possíveis que reduziriam a probabilidade da hipótese a quase zero; deve haver fenômenos que a hipótese não consegue explicar, resultados experimentais concebíveis que a teoria não pode acomodar. Uma teoria que pode explicar tudo e não proíbe nada, não fornece orientação sobre o que esperar.

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_j P(B|A_j)P(A_j)}$$

Em termos do Teorema de Bayes, se houver pelo menos uma observação B que a hipótese A_i não consegue explicar, ou seja, $P(B|A_i)$ é extremamente baixa, então o numerador $P(B|A_i)\ P(A_i)$ também será ínfimo, e, portanto, a probabilidade posterior $P(A_i|B)$ também será extremamente baixa. A observação de um resultado improvável B reduziria a probabilidade de A_i a quase zero. Uma teoria que se recusa a se expor a essa vulnerabilidade precisa distribuir sua probabilidade amplamente para evitar lacunas; não pode concentrar fortemente a probabilidade em alguns resultados preferidos. não será capaz de oferecer orientações precisas.

Essa é a regra científica derivada da teoria da probabilidade.

Dentro desse contexto, ‘falsificabilidade’ é algo que você avalia analisando uma única hipótese e perguntando: “Quão estreitamente ela concentra sua distribuição de probabilidades sobre os resultados possíveis? Até que ponto ela me diz o que esperar? Ela consegue explicar alguns resultados muito melhor do que outros?”

A interpretação da decoerência da mecânica quântica é falsificável? Existem resultados experimentais que poderiam reduzir sua probabilidade a quase zero?

Claro que sim. Poderíamos medir partículas emaranhadas que deveriam sempre ter spin opostos e descobrir que, quando medidas a uma distância suficiente, às vezes têm o mesmo spin.

Ou poderíamos observar maçãs caindo para cima, os planetas do Sistema Solar movendo-se aleatoriamente e um átomo emitindo fótons sem uma fonte de energia aparente. Essas observações também falsificariam a interpretação decoerente da mecânica quântica. São eventos que não deveríamos esperar ocorrer sob a suposição de que a decoerência governa o universo.

Portanto, existem observações B para as quais $P(B|A_{deco})$ é praticamente zero, levando a $P(A_{deco}|B)$ também ser praticamente zero.

Mas isso ocorre porque a decoerência ainda é uma forma de mecânica quântica! E quanto à parte da decoerência em comparação com o postulado do colapso?

Estamos chegando lá. O ponto é que acabei de definir um teste que nos faz considerar uma hipótese de cada vez (chamando-a de ‘falsificabilidade’). Se desejamos distinguir a decoerência do colapso, precisamos considerar pelo menos duas hipóteses ao mesmo tempo.

Na verdade, o teste de ‘falsificabilidade’ não é tão focado em uma única hipótese, já que a soma no denominador deve envolver alguma outra hipótese. No entanto, o que defini como ‘falsificabilidade’ expõe o tipo de problema que Karl Popper estava apontando quando afirmou que a psicanálise freudiana era ‘infalsificável’ porque podia explicar qualquer comportamento humano sem restrições.

Se fôssemos uma espécie alienígena sem familiaridade com o postulado do colapso ou a Interpretação de Copenhague, e nossa única teoria física fosse limitada à mecânica quântica decoerente - se tudo que você tivesse na cabeça fossea equação diferencial para a evolução da função de onda mais a regra da probabilidade de Born - você ainda teria expectativas nítidas em relação ao universo. Você não viveria em um mundo mágico onde tudo é provável.

Mas a mesma afirmação poderia ser feita em relação à mecânica quântica sem a decoerência (no nível macroscópico).

É verdade! Alguém que anda por aí com a equação para a evolução da função de onda, além do postulado de colapso que obedece às probabilidades de Born e é acionado antes que a superposição atinja níveis macroscópicos, ainda vive em um universo onde as maçãs caem em vez de subir.

Mas onde a decoerência faz uma nova previsão, uma que possa ser testada?

Uma “nova” previsão em relação a quê? Em relação ao conhecimento da época dos antigos gregos? Se voltássemos no tempo e mostrássemos a eles a mecânica quântica decoerente, eles poderiam fazer previsões experimentais que antes eram impossíveis.

Quando falamos em ‘nova previsão’, estamos nos referindo a algo ‘novo’ em relação a alguma outra hipótese que define a ‘antiga previsão’. Isso nos conduz à teoria do que optei por chamar de testabilidade; e o algoritmo considera, inerentemente, pelo menos duas hipóteses ao mesmo tempo. Não podemos denominar algo como ‘nova previsão’ baseando-nos apenas em uma hipótese isolada.

Em termos bayesianos, buscamos um elemento de evidência B que fornecerá evidências a favor de uma hipótese em detrimento de outra, diferenciando-as, e o processo de geração dessa evidência poderíamos chamar de ‘teste’. Estamos procurando um resultado experimental B tal que

$$P(B|A_d) \neq P(B|A_c);$$

ou seja, algum resultado B que tenha uma probabilidade diferente, condicionada à hipótese de decoerência ser verdadeira, em comparação com a probabilidade se a hipótese de colapso for verdadeira. Isso, por sua vez, implica que as probabilidades posteriores de decoerência e colapso se tornarão distintas das probabilidades anteriores:

$$\frac{P(B|A_d)}{P(B|A_c)} \neq 1$$

implica que

$$\frac{P(A_d|B)}{P(A_c|B)} = \frac{P(B|A_d)}{P(B|A_c)} \times \frac{P(A_d)}{P(A_c)}$$

$$\frac{P(A_d|B)}{P(A_c|B)} \neq \frac{P(A_d)}{P(A_c)}.$$

Esta equação é simétrica (assumindo que nenhuma probabilidade seja literalmente igual a 0). Não há um A_j rotulado como 'antiga hipótese' e outro A_j rotulado como 'nova hipótese'.

Essa simetria é uma característica, e não um defeito, da teoria da probabilidade! Se você estiver projetando um sistema de raciocínio artificial que chegue a diferentes crenças, dependendo da ordem na qual as evidências são apresentadas, isso é chamado de 'histerese' e é considerado algo negativo. Ouvi dizer que também é desaprovado na ciência.

Do ponto de vista da teoria da probabilidade, existem vários teoremas triviais que dizem que não importa se você atualiza X primeiro e depois Y, ou se atualiza Y primeiro, depois X. Pelo menos eles seriam triviais se os seres humanos não os violassem com tanta frequência e de forma tão leviana.

Se a decoerência é 'instável' em relação ao colapso, então o colapso também é 'instável' em relação à decoerência. E se a história da física tivesse seguido um caminho diferente - e se Hugh Everett e John Wheeler tivessem ocupado o lugar de Bohr e Heisenberg, e vice-versa? Seria correto e apropriado que as pessoas deste mundo olhassem para a interpretação do colapso, franzissem as sobrancelhas e perguntassem: 'Onde estão as novas previsões?'

E se, algum dia, encontrarmos uma espécie alienígena que tenha descoberto a decoerência antes do colapso? Cada um de nós estaria obrigado a manter a teoria que concebemos primeiro? A razão não teria nada a dizer sobre o assunto, deixando-nos com nenhum recurso para resolver o argumento, a não ser a guerra interestelar?

Mas, se abandonarmos a exigência de produzir novas previsões, ficaremos no caos científico. Poderíamos acrescentar complicações arbitrárias e não testáveis às teorias existentes e obter previsões experimentalmente equivalentes.

Mas se revogarmos a exigência de produzir novas previsões, teremos o caos científico. Você pode adicionar complicações arbitrárias e não testáveis a teorias antigas e obter previsões experimentalmente equivalentes. Se rejeitarmos o que você chama de 'histerese', como poderemos defender nossas teorias atuais contra todos os lunáticos que propõem que os elétrons possuam uma nova propriedade chamada 'cheiro', assim como os quarks possuem 'sabor'?

Primeiramente, devo concordar que não devemos aceitar alguém que chegue até nós e diga: 'Ei, te-

nho uma ideia brilhante! Talvez o campo eletromagnético não esteja atraindo as partículas carregadas. Talvez existam pequenos anjos que, na verdade, empurram as partículas, e o campo eletromagnético apenas lhes diz como fazer isso. Olha, tenho todas essas previsões experimentais bem-sucedidas - as previsões que você costumava considerar suas!

Portanto, concordo que não devemos adotar essa nova teoria incrível, mas o problema não é a novidade em si.

Suponhamos que a história da humanidade tenha se desenrolado apenas ligeiramente de forma diferente, com a Igreja sendo a principal entidade financiadora da Ciência. E suponhamos que, quando as leis do eletromagnetismo foram formuladas pela primeira vez, o fenômeno do magnetismo tenha sido considerado evidência da existência de espíritos invisíveis, de anjos. James Clerk Maxwell tornou-se São Maxwell, aquele que descreveu as leis que governam as ações dos anjos.

Alguns séculos depois, quando o poder da Igreja de queimar pessoas na fogueira estava diminuindo, alguém chega e diz: 'Ei, nós realmente precisamos dos anjos?'

'Sim', todos respondem. 'Como, de outra forma, os simples números do campo eletromagnético se traduziriam nos movimentos reais das partículas?'

'Pode ser uma lei fundamental', diz o recém-chegado, 'ou pode ser algo diferente dos anjos, que descobriremos mais tarde. O que estou sugerindo é que interpretar os números como a ação dos anjos não adiciona nada, e devemos manter apenas os números e descartar a parte dos anjos.'

Eles se olham e finalmente dizem: "Mas sua teoria não faz nenhuma nova previsão experimental, então por que deveríamos adotá-la? Como podemos testar suas alegações sobre a ausência de anjos?"

Do ponto de vista normativo, parece-me que, se devemos rejeitar os anjos insanos no primeiro cenário, mesmo sem conseguirmos distinguir experimentalmente as duas teorias, então também devemos rejeitar os anjos da ciência estabelecida no segundo cenário, mesmo sem conseguirmos distinguir experimentalmente as duas teorias.

Normalmente, é o lunático que acrescenta complicações inúteis, em vez de cientistas que as introduzem acidentalmente no início. Mas o problema não é que as complicações sejam novas, mas sim que sejam inúteis, quer sejam novas ou não.

Um bayesiano diria que as complicações adicionais dos anjos na teoria resultam em penalidades na probabilidade a priori da teoria. Se duas teorias fazem previsões equivalentes, devemos manter aquela que pode ser descrita na mensagem mais curta, o programa mais simples. Se você estiver avaliando a probabilidade a priori de cada hipótese contando os bits de código e, em seguida, aplicando as regras de atualização bayesiana a todas as evidências disponíveis, não fará diferença qual hipótese você ouve primeiro ou a ordem em que aplica as evidências.

Geralmente, não é possível aplicar a teoria da probabilidade formal na realidade, assim como não é possível prever o vencedor de uma partida de tênis usando a teoria quântica de campos. Mas, se a teoria da probabilidade pode servir como um guia para a prática, é isso que ela diz: rejeite as complicações inúteis, em geral, não apenas quando são novas.

Sim, e inúteis é exatamente o que são os muitos mundos da decoerência! Supostamente, todos esses mundos existem ao lado do nosso, e eles não têm impacto em nosso mundo, mas mesmo assim, devo acreditar neles?

Não, segundo a decoerência, o que você deve acreditar são as leis gerais que governam as funções de onda - e essas leis gerais são muito visíveis e testáveis.

Argumentei em outro lugar que o selo da ciência deve estar associado a leis gerais e não a eventos particulares, porque são as leis gerais que, em princípio, qualquer pessoa pode testar por si mesma. Garanto a você que estou usando meias brancas agora enquanto digito isso. Portanto, você provavelmente está racionalmen-

te justificado em acreditar que este é um fato histórico. Contudo, não é o tipo de afirmação especialmente forte que canonizamos como uma crença provisória da ciência, porque não há nenhum experimento que você possa realizar por si mesmo para determinar a verdade dela; você está confiando em minha autoridade. Agora, se eu dissesse a você a massa de um elétron em geral, você poderia sair e encontrar seu próprio elétron para testar e, assim, verificar você mesmo a veracidade da lei geral naquele caso específico.

A capacidade de qualquer pessoa sair e verificar uma lei científica geral por si mesma, construindo algum caso particular, é o que torna nossa crença na lei geral especialmente confiável.

O que os decoerentistas afirmam acreditar é na equação diferencial que observamos governar a evolução das funções de onda - e essa equação pode ser testada a qualquer momento; basta olhar para um átomo de hidrogênio.

A crença na existência de partes separadas da função de onda universal não é um acréscimo e não deve explicar o preço do ouro em Londres; é simplesmente uma consequência dedutiva da evolução da função de onda. Se a evidência de muitos casos particulares lhe dá razões para acreditar que X→Y é uma lei geral, e a evidência de algum caso particular lhe dá razões para acreditar em X, então você deve ter P(Y) ≥ P(X e (X→Y)).

Ou, olhando de outra forma, se

P(Y|X) ≈ 1, então P(X e Y) ≈ P(X).

Isso significa que acreditar em detalhes adicionais não lhe custa uma probabilidade adicional quando eles são implicações lógicas de crenças gerais que você já possui. Presumivelmente, as próprias crenças gerais são falsificáveis, ou por que se preocupar?

É por isso que não acreditamos que as naves desapareçam quando cruzam o horizonte cosmológico relativo a nós. É verdade que a existência contínua das espaçonaves não afeta nosso mundo. A existência contínua das espaçonaves não ajuda a explicar o preço do ouro em Londres. No entanto, obtemos a espaçonave invisível sem custo adicional, como uma consequência das leis gerais que implicam a conservação de massa e energia. Se a existência contínua das espaçonaves não fosse uma consequência dedutiva das leis da física, conforme as modelamos atualmente, então seria um detalhe adicional, custaria uma probabilidade extra e teríamos que nos perguntar por que nossa teoria deve incluir essa afirmação.

A parte da decoerência que deve ser testada não são os muitos mundos em si, mas apenas a lei geral que governa a função de onda. Os decoerentistas observam que, aplicada universalmente, essa lei implica a existência de mundos inteiros sobrepostos. Agora, existem críticas que podem ser feitas a essa teoria, mais notavelmente, “Mas então de onde vêm as probabilidades de Born?” Mas dentro da lógica interna da decoerência, os muitos mundos não são oferecidos como explicação para nada, nem são a substância da teoria que deve ser testada; eles são simplesmente uma consequência lógica das leis gerais que constituem a substância da teoria.Se A ⇒ B então ¬B ⇒ ¬A. Negar a existência de mundos superpostos é necessariamente negar a universalidade das leis quânticas formuladas para governar os átomos de hidrogênio e todos os outros casos passíveis de exame; é essa negação que parece aos decoerentistas o detalhe extra e não testável. Você não pode ver as outras partes da função de onda - por que postular adicionalmente que elas não existem?

Os eventos em torno da controvérsia da decoerência podem ser únicos na história científica, marcando a primeira vez que cientistas sérios se apresentaram e disseram que, por acidente histórico, a humanidade desenvolveu uma teoria física matemática poderosa e bem-sucedida que inclui anjos. Que existe toda uma lei, o postulado do colapso, que pode simplesmente ser jogado fora, deixando a teoria estritamente mais simples.

Para esta discussão, desejo contribuir com a afirmação de que, à luz de uma compreensão matematicamente sólida da teoria da probabilidade, a decoerência não é descartada pela Navalha de Ocam, nem é infalsificável, nem não é testável.

Podemos considerar, por exemplo, a decoerência e o postulado do colapso, lado a lado, e avaliar

críticas como “A decoerência definitivamente não prevê que as probabilidades quânticas devem ser sempre 50/50?” e “O colapso não viola a Relatividade Especial ao implicar influência à distância?” Podemos considerar os méritos relativos dessas teorias com base em sua compatibilidade com a experiência e o caráter aparente da lei física.

Afirmar que a decoerência nem sequer está no jogo - porque os muitos mundos em si são “entidades adicionais” que violam a Navalha de Ocam, ou porque os próprios muitos mundos são “intestáveis”, ou porque a decoerência não faz “novas previsões” - tudo isso é, eu argumentaria, um erro absoluto da teoria da probabilidade. Esses argumentos específicos devem ser simplesmente descartados e a discussão seguir em frente.

236 - Privilegiando a hipótese

Imagine que a polícia de Cidadona, uma cidade com um milhão de habitantes, está investigando um assassinato com poucas ou nenhuma pista. A vítima foi esfaqueada em um beco, sem impressões digitais ou testemunhas.

Então, um dos detetives diz:

"Bem... não temos ideia de quem fez isso... nenhuma evidência específica que aponte para qualquer uma das milhões de pessoas nesta cidade... mas vamos considerar a hipótese de que este assassinato foi cometido por Mortimer Q. Snodgrass, que mora na Rua Comum, 128. Poderia ter sido ele, afinal."

(Chamarei isso de falácia de privilegiar a hipótese. (Me avisem se já existe um nome oficial - não me lembro de tê-la visto descrita antes.)

O detetive pode ter alguma forma de evidência racional que não seja admissível no tribunal - como um boato de um informante, por exemplo. Mas se o detetive não tiver alguma justificativa prévia para promover Mortimer à atenção especial da polícia - se o nome foi tirado completamente do nada - então os direitos de Mortimer estão sendo violados.

E isso é verdade, mesmo que o detetive não esteja afirmando que Mortimer "realmente" fez isso, mas apenas pedindo à polícia para considerar que Mortimer possa ter feito. Está promovendo injustificadamente essa hipótese específica à atenção. É da natureza humana procurar confirmação em vez de refutação. Suponha que três detetives sugiram seus inimigos como nomes a serem considerados: Mortimer tem cabelos castanhos, Frederico tem cabelos pretos e Helena é loira. Então, uma testemunha é encontrada e diz que a pessoa saindo da cena tinha cabelos castanhos. "Ahá!", dizem os policiais. "Antes não tínhamos evidências para distinguir entre as possibilidades, mas agora sabemos que foi Mortimer!"

Isso está relacionado ao princípio que comecei a chamar de "localizar a hipótese". Se você tem um bilhão de caixas e apenas uma contém um diamante (a verdade), e seus detectores fornecem apenas 1 bit de evidência cada, é preciso muito mais evidência para promover a verdade à sua atenção particular - para reduzi-la a dez boas possibilidades, cada uma merecendo nossa atenção individual - do que para descobrir qual dessas dez possibilidades é verdadeira. São necessários 27 bits para reduzi-la a dez, e apenas mais 4 bits nos darão chances melhores que 50% de ter a resposta certa.

Assim, o detetive, ao chamar a atenção da polícia para Mortimer, sem motivo, entre um milhão de outras pessoas, está pulando a maioria das evidências que precisam ser apresentadas contra Mortimer.

E o detetive deve ter essas evidências em mãos no primeiro momento em que trouxer Mortimer à atenção da polícia. Pode ser mera evidência racional e não legal, mas se não há evidência, o detetive está assediando e perseguindo o pobre Mortimer.

Durante meu recente diávlogo com Scott Aaronson sobre mecânica quântica, consegui encurralar Scott ao ponto de fazê-lo admitir que não havia evidência concreta favorecendo um postulado de colapso ou mecânica quântica de mundo único. Mas, disse Scott, poderíamos encontrar evidências futuras a favor da mecânica quântica de mundo único, e a teoria de muitos mundos ainda tem a questão em aberto das probabilidades de Born.

Isso é exatamente o que eu chamaria de falácia de privilegiar a hipótese. Devem existir um trilhão de maneiras melhores de responder a questão do Born sem adicionar um postulado de colapso que é a única lei não-linear, não-unitária, descontínua, não-diferenciável, não-simétrica-CPT, não-local no espaço de configuração, que viola o Teorema de Liouville, tem um espaço privilegiado de simultaneidade, influencia mais rápido que a luz, não é causal e é uma lei informalmente especificada em toda a física. Algo tão não-físico é inútil ser dito em voz alta ou mesmo pensado como possibilidade sem um peso bastante grande de evidências - muito mais do que o atual total de zero.

Mas devido a um acidente histórico, postulados de colapso e mecânica quântica de mundo único estão de fato nos lábios e mentes de todos para serem pensados. Assim, a questão em aberto das probabilidades de Born é oferecida (por ninguém menos que Scott Aaronson!) como evidência de que a teoria de muitos mundos ainda não pode oferecer uma imagem completa do mundo. Isso é interpretado como se a mecânica quântica de mundo único ainda estivesse de alguma forma na disputa.

Na mente das pessoas, se você consegue fazê-las pensar sobre essa hipótese particular em vez das trilhões de outras possibilidades que não são mais complicadas ou improváveis, você realmente fez uma grande parte do trabalho de persuasão. Qualquer coisa pensada é tratada como "na disputa", e se outros competidores parecem ficar um pouco para trás na corrida, assume-se que este competidor está avançando ou até mesmo assumindo a liderança.

E sim, esta é a mesma falácia cometida, em uma escala muito mais flagrante, pelo teísta que aponta que a ciência moderna não oferece uma explicação absolutamente completa de todo o universo, e considera isso como evidência da existência de Jeová. Em vez de Alá, o Monstro do Espaguete Voador, ou trilhões de outros deuses não menos complicados - sem falar no espaço de explicações naturalistas!

Falar sobre "projeto inteligente" sempre que você aponta uma suposta falha ou problema em aberto na teoria evolutiva é, novamente, privilegiar a hipótese. Você já deve ter evidências em mãos que apontem especificamente para o projeto inteligente para justificar levantar essa ideia particular à nossa atenção, em vez de milhares de outras.

Então, essa é a regra sensata. E a anti-epistemologia correspondente é falar infinitamente sobre "possibilidade" e como você "não pode refutar" uma ideia, esperar que evidências futuras possam confirmá-la sem apresentar evidências passadas já em mãos, focar e focar em possibilidades sem avaliar evidências possivelmente desfavoráveis, desenhar imagens verbais brilhantes de observações confirmatórias que poderiam acontecer, mas não aconteceram, ou tentar mostrar que pedaço após pedaço de evidência negativa "não é conclusivo".

Da mesma maneira que a Navalha de Occam diz que proposições mais complicadas requerem mais evidências para serem acreditadas, proposições mais complicadas também deveriam exigir mais trabalho para serem levantadas à atenção. Assim como o princípio dos detalhes onerosos exige que cada parte de uma crença seja justificada separadamente, ele exige que cada parte seja levantada à atenção separadamente.

Como discutido em Crenças de Movimento Perpétuo, a fé e as máquinas de movimento perpétuo tipo 2 (água→cubos de gelo + eletricidade) têm em comum que pretendem fabricar improbabilidade do nada, seja a improbabilidade da água formar cubos de gelo ou a improbabilidade de chegar a crenças corretas sem observação. Às vezes, a maioria do anti-trabalho envolvido na fabricação dessa improbabilidade é nos fazer prestar atenção a uma crença injustificada - pensando nela, refletindo sobre ela. Em grandes espaços de respostas, atenção sem evidência é mais da metade do caminho para a crença sem evidência.

Alguém que passa o dia todo pensando se a Trindade existe ou não, em vez de Alá ou Thor, ou o Monstro do Espaguete Voador, está mais da metade do caminho para o Cristianismo. Se estiver saindo, está menos da metade partido; se estiver chegando, está mais da metade lá.

Um modo de privilégio encontrado frequentemente é tentar fazer a incerteza num espaço transbordar para fora desse espaço sobre a hipótese privilegiada. Por exemplo, um criacionista se agarra a algum aspecto (supostamente) debatido da teoria contemporânea, argumenta que os cientistas estão incertos sobre a evolução e então diz: "Não sabemos realmente qual teoria está certa, então talvez o projeto inteligente

esteja certo.” Mas a incerteza é incerteza no reino das teorias naturalistas da evolução - não temos razão para acreditar que precisaremos sair desse reino para lidar com nossa incerteza, muito menos que saltaríamos para fora do reino da ciência padrão e cairíamos em Jeová em particular. Isso é privilegiar a hipótese - pegar a dúvida em um espaço normal e tentar fazer a dúvida transbordar para fora do espaço normal, sobre um alvo privilegiado (e geralmente desacreditado) extremamente anormal.

Da mesma forma, nossa incerteza sobre a origem das estatísticas de Born deveria ser incerteza no espaço de teorias quânticas que são contínuas, lineares, unitárias, mais lentas que a luz, locais, causais, naturalistas, etc. - o caráter usual da lei física. Parte dessa incerteza pode transbordar para fora do espaço padrão em teorias que violam uma dessas características padrão. É de fato possível que possamos ter que pensar fora da caixa. Mas teorias de mundo único violam todas essas características, e não há razão para privilegiar essa hipótese.”

237 - Vivendo em muitos mundos

Alguns comentaristas recentemente expressaram perturbação com a ideia de se dividir constantemente em zilhões de outras pessoas, como é a previsão direta e inevitável da mecânica quântica.

Outros confessaram não entender as implicações dos muitos mundos para o planejamento. Se você decide apertar o cinto de segurança neste mundo, isso aumenta a chance de outro eu desapertar o cinto? Você está sendo egoísta às custas deles?

Apenas se lembre da Lei de Egan: Tudo se soma à normalidade.

(Citando Greg Egan, em *Quarentine* (Quarentena). [1])

Frank Sulloway disse: [2]

> Ironicamente, a psicanálise supera o darwinismo precisamente porque suas previsões são tão estranhas e suas explicações tão contra-intuitivas que pensamos: 'Isso é realmente verdade? Que radical!' As ideias de Freud são tão intrigantes que as pessoas estão dispostas a pagar por elas, enquanto uma das grandes desvantagens do darwinismo é que sentimos que já o conhecemos, porque, de certo modo, conhecemos[46].

Quando Einstein derrubou a versão newtoniana da gravidade, as maçãs não pararam de cair e os planetas não se desviaram para o Sol. Toda nova teoria da física deve capturar as previsões bem-sucedidas da antiga teoria que substituiu. Ela deve prever que o céu será azul, não verde.

Então não pense que muitos mundos estão aí para fazer previsões estranhas, radicais e empolgantes. Tudo se soma à normalidade.

Então por que alguém deveria se importar?

Porque uma vez foi feita a pergunta, fascinante para um racionalista: Qual tudo soma à normalidade?

E a resposta a esta pergunta acaba sendo: mecânica quântica. É a mecânica quântica que se soma à normalidade.

Se houvesse algo diferente no lugar da mecânica quântica, então o mundo pareceria estranho e incomum.

Tenha isso em mente quando estiver se perguntando como viver no novo e estranho universo de muitos mundos: você sempre esteve lá.

As religiões, dizem os antropólogos, geralmente exibem uma propriedade chamada contra-intuitividade mínima. Elas são surpreendentes o suficiente para serem memoráveis, mas não tão bizarras a ponto de serem difíceis de memorizar. Anúbis tem a cabeça de um cão, o que o torna memorável, mas o resto dele é o

46 NT. Texto original em inglês. *Ironically, psychoanalysis has it over Darwinism precisely because its predictions are so outlandish and its explanations are so counterintuitive that we think, Is that really true? How radical! Freud's ideas are so intriguing that people are willing to pay for them, while one of the great disadvantages of Darwinism is that we feel we know it already, because, in a sense, we do.*

corpo de um homem. Espíritos podem ver através das paredes, mas ainda ficam com fome.

Mas a física não é uma religião, feita para surpreendê-lo apenas o suficiente para ser memorável. Os fenômenos subjacentes são tão contra-intuitivos que é preciso um longo estudo para que os humanos os compreendam. Mas os fenômenos superficiais são totalmente comuns. Você nunca verá outro mundo com o canto do olho. Você nunca ouvirá a voz de algum outro eu. Isso é inequivocamente proibido pelas leis. Desculpe, você é apenas esquizofrênico.

O ato de tomar decisões não tem interação especial com o processo que ramifica mundos. Em sua mente, em sua imaginação, uma decisão parece um ponto de ramificação onde o mundo poderia seguir dois caminhos diferentes. Mas você sentiria a mesma incerteza, visualizaria as mesmas alternativas, se houvesse apenas um mundo. É o que as pessoas pensaram por séculos antes da mecânica quântica, e ainda visualizavam resultados alternativos que poderiam resultar de suas decisões.

Decisão e decoerência são conceitos totalmente ortogonais. Se seu cérebro nunca se tornasse decoerente, esse único processo cognitivo ainda teria que imaginar escolhas diferentes e seus diferentes resultados. E uma pedra, que não toma decisões, obedece às mesmas leis da mecânica quântica que qualquer outra coisa, e se divide freneticamente enquanto permanece em um lugar.

Você não se divide particularmente quando chega a uma decisão, assim como não se divide particularmente quando respira. Você está apenas se dividindo o tempo todo como resultado da decoerência, que não tem nada a ver com escolhas.

Existe uma população de mundos, e em cada mundo, tudo se soma à normalidade: as maçãs não param de cair. Em cada mundo, as pessoas escolhem o curso que lhes parece melhor. Talvez elas sigam uma linha de pensamento diferente, vejam novas implicações ou percam outras, e cheguem a uma escolha diferente. Mas não é que um mundo escolha cada opção. Não é que uma versão de você escolha o que parece melhor, e outra versão escolha o que parece pior. Em cada mundo, as maçãs continuam caindo e as pessoas continuam fazendo o que parece uma boa ideia.

Sim, você pode encontrar exceções a essa regra, mas são exceções normais. Tudo se soma à normalidade, em todos os mundos.

Você não pode “escolher em qual mundo acabar”. Em todos os mundos, as escolhas das pessoas determinam os resultados da mesma forma que determinariam em apenas um único mundo.

A escolha que você faz aqui não tem alguma influência equilibradora estranha em algum mundo em outro lugar. Não há comunicação causal entre mundos decoerentes. Em cada mundo, as escolhas das pessoas controlam o futuro daquele mundo, não de algum outro mundo.

Se você pode imaginar a tomada de decisões em um mundo, pode imaginar a tomada de decisões em muitos mundos: basta ter o mundo se dividindo constantemente enquanto obedece a todas as mesmas regras.

Em nenhum mundo dois mais dois é igual a cinco. Em nenhum mundo as naves espaciais podem viajar mais rápido que a luz. Todos os mundos quânticos obedecem às nossas leis da física; sua existência é afirmada em primeiro lugar por nossas leis da física. Desde o início, nada incomum jamais aconteceu, neste ou em qualquer outro mundo. Todos são legais.

Existem mundos horríveis por aí, totalmente além da sua capacidade de afetar? Claro. E coisas horríveis aconteceram durante o século XII, que também estão além da sua capacidade de afetar. Mas o século XII não é sua responsabilidade, porque, como diz a frase antiquada, “já aconteceu”. Eu sugeriria que você considere cada mundo que não está em seu futuro como parte do “passado generalizado”.

Viva em seu próprio mundo. Antes de saber sobre física quântica, você não teria sido tentado a tentar viver em um mundo que não parecia existir. Suas decisões devem se somar a essa mesma normalidade: você não deve tentar viver em um mundo quântico com o qual não pode se comunicar.

Sua teoria de tomada de decisões deve (quase sempre) ser a mesma, quer você suponha haver 90% de probabilidade de algo acontecer, ou se acontecerá em 9 de 10 mundos. Agora, como as pessoas têm dificuldade em lidar com probabilidades, pode ser útil visualizar algo acontecendo em 9 de 10 mundos. Mas isso apenas ajuda você a usar a teoria normal de tomada de decisões.

Agora é um bom momento para começar a aprender a calar a boca e multiplicar. Como observo em "Loterias: Um Desperdício de Esperança":

"O cérebro humano não faz aritmética de ponto flutuante de 64 bits, e não pode desvalorizar a força emocional de uma antecipação agradável por um fator de 0,00000001 sem abandonar completamente a linha de raciocínio."

E em "Nova Loteria Melhorada":

"Entre zero chance de ficar rico e épsilon chance, há uma diferença da ordem de épsilon. Se você duvida disso, deixe épsilon igual a um sobre googolplex."

Se você está pensando em um mundo que poderia surgir de maneira legal, mas cuja probabilidade é de um quadrilhão para um, e algo muito agradável ou muito terrível está acontecendo neste mundo… bem, provavelmente existe, se for legal. Mas você deve tentar liberar um quadrilionésimo de neurotransmissores, em seus centros de recompensa ou aversão, para poder pesar esse mundo adequadamente em suas decisões. Se você não acha que pode fazer isso… não se incomode em pensar nisso.

Caso contrário, você pode muito bem sair e comprar um bilhete de loteria usando um número aleatório quântico, uma estratégia garantida para resultar em uma mega-vitória muito pequena.

Ou aqui está outra maneira de pensar sobre isso: você está considerando gastar alguma energia mental em um mundo cuja frequência em seu futuro é menor que um trilionésimo? Então vá buscar um dado de 10 lados na sua loja de jogos local e, antes de começar a pensar naquele mundo estranho, comece a rolar o dado. Se o dado cair em 9 doze vezes seguidas, então você pode pensar naquele mundo. Caso contrário, não perca seu tempo; o tempo de pensamento é um recurso a ser gasto com sabedoria.

Você pode rolar os dados quantas vezes quiser, mas não pode pensar no mundo até que 9 apareça doze vezes seguidas. Então você pode pensar nisso por um minuto. Depois disso, você tem que começar a rolar o dado novamente.

Isso pode ajudá-lo a apreciar o conceito de "um trilhão para um" em um nível mais visceral.

Se em algum momento você se pegar pensando que a física quântica pode ter algum tipo de implicação estranha e anormal para a vida cotidiana - então você provavelmente deve parar por aí.

Oh, há algumas implicações de muitos mundos para a ética. O utilitarismo médio de repente parece muito mais atraente - você não precisa se preocupar em criar o maior número possível de pessoas, porque já há muitas pessoas explorando o espaço-pessoa. Você só quer que a qualidade média de vida seja a mais alta possível, nos mundos futuros que são sua responsabilidade.

E você sempre deve se alegrar com a descoberta, desde que você pessoalmente não saiba uma coisa. É sem sentido falar em ser a "primeira" ou a "única" pessoa a saber uma coisa, quando tudo o que é conhecível é conhecido em mundos que não estão nem em seu passado, nem em seu futuro, e não estão nem antes, nem depois de você.

Mas, em geral, tudo se soma à normalidade. Se sua compreensão de muitos mundos é minimamente instável, e você está contemplando se deve acreditar em alguma proposição estranha, sentir alguma emoção estranha ou planejar alguma estratégia estranha, então posso lhe dar um conselho muito simples: não faça isso.

O universo quântico não é um lugar estranho para o qual você foi empurrado. É assim que as coisas sempre foram.

Referências

[1] Greg Egan, Quarantine (London: Legend Press, 1992).

[2] Robert S. Boynton, “The Birth of an Idea: A Profile of Frank Sulloway,” The New Yorker (October 1999).

238 - Não-realismo quântico

A lua existe quando ninguém a observa? [47]

— Albert Einstein, perguntado a Niels Bohr

Suponha que você esteja no início de uma jornada para desenvolver uma teoria da mecânica quântica.

À medida que avança, encontra experimentos que produzem resultados distintos, dependendo do nível de observação. Você se aprofunda na realidade conhecida e encontra uma descrição matemática incrivelmente precisa que fornece apenas a frequência relativa dos resultados; pior ainda, essa descrição é composta por números complexos. As coisas se comportam como partículas na segunda-feira e como ondas na terça-feira.

A resposta correta não está disponível para você como uma hipótese, porque só será concebida daqui a trinta anos.

Diante dessa perplexidade, qual é a melhor abordagem?

A melhor estratégia é adotar a rigorosa interpretação da "cale a boca e calcule" da mecânica quântica. Você continuará buscando o desenvolvimento de novas teorias, pois fazer o melhor que puder não significa desistir. Mas especificamos que a resposta correta não estará disponível por trinta anos, e isso significa que nenhuma das novas teorias será realmente boa. Fazer o melhor que você teoricamente pode, envolve reconhecer essa limitação, mesmo quando você procurou maneiras de testar as hipóteses.

Fazer o melhor sob tais circunstâncias não incluiria declarar algo como: "A função de onda fornece apenas probabilidades, não certezas." Essa afirmação, à luz da retrospectiva, foi uma conclusão precipitada; a função de onda garante a existência de múltiplos mundos. Portanto, a ideia de que a função de onda é meramente uma probabilidade estava equivocada. Você realizou cálculos, mas não conseguiu permanecer em silêncio.

Se você se esforçar para fazer o melhor quando a resposta correta não está disponível, então quando ouvir falar em decoerência, verá que não disse nada incompatível com a decoerência. A decoerência não é descartada pelos dados e cálculos. Portanto, ao se recusar a afirmar, como conhecimento positivo, qualquer proposição que não tenha sido forçada por dados e cálculos, os cálculos não o forçarão a dizer algo incompatível com a decoerência. Isso se aplica igualmente a qualquer que seja a teoria correta, se ela não for a decoerência. Se você se desviar do caminho, deve ser por iniciativa própria.

Entretanto, é desafiador para os seres humanos manterem-se em silêncio e calcular - verdadeiramente se calar e calcular. Há uma tendência avassaladora de tratar nossa ignorância como se fosse conhecimento positivo.

Não posso confirmar se ocorreu alguma conversa exatamente assim, mas é assim que a ignorância é

47 NT. Texto original em inglês. *Does the moon exist when no one is looking at it?*

frequentemente transformada em conhecimento:

GALLANT: “Cale a boca e calcule.”

GOOFUS: “Por quê?”

GALLANT: “Porque não sei o que essas equações significam, apenas sei que funcionam.”

Cinco minutos depois -

GOOFUS: “Cale a boca e calcule.”

ALUNO: “Por quê?”

GOOFUS: “Porque essas equações não têm significado, apenas funcionam.”

ALUNO: “Mesmo? Como você sabe?”

GOOFUS: “Gallant me disse.”

Uma transformação semelhante ocorre quando passamos de:

GALLANT: “Quando meus cálculos mostram uma amplitude de (−1/3)i para a absorção deste fóton, meus experimentos indicam que o fóton foi absorvido cerca de 107 vezes em 1.000, o que se alinha com 1/9, o quadrado do módulo. Há, claramente, uma relação entre as estatísticas experimentais e o módulo ao quadrado da amplitude, mas não sei qual é.

GOOFUS: “A amplitude de probabilidade não determina a localização do elétron, apenas onde ele pode estar. O quadrado do módulo é a probabilidade de a realidade se manifestar dessa forma. A própria realidade é inerentemente não determinística.”

E novamente:

GALLANT: “Após medir algo e obter um resultado experimental, faço cálculos futuros usando apenas a amplitude cujo módulo ao quadrado foi usado para calcular a frequência desse resultado experimental. Seguir essa regra garante que meus cálculos subsequentes coincidam com as frequências observadas.”

GOOFUS: “Uma vez que a amplitude representa probabilidade, após obter um resultado experimental, a probabilidade de todas as outras possibilidades torna-se zero!”

A mudança completa de:

O quadrado dessa “amplitude” corresponde precisamente às nossas frequências observadas experimentalmente

para

A amplitude representa a probabilidade de um resultado de medição

para

Claro, uma vez que você sabe que não obteve uma medição, a probabilidade se torna zero.

Esse deve ser um dos erros mais embaraçosos na história da ciência.

Se levarmos isso literalmente, teríamos a interpretação da consciência como causa do colapso na mecânica quântica. Hoje em dia, quase ninguém admitiria realmente acreditar de fato na interpretação da mecânica quântica que a consciência causa o colapso.

Mas os livros de física ainda são escritos dessa maneira. As pessoas dizem que não acreditam, mas

falam como se o conhecimento estivesse eliminando amplitudes de “probabilidade” incompatíveis.

Contudo, por mais improvável que pareça a interpretação da consciência-causa-colapso, ela pelo menos nos fornece uma imagem da realidade. É uma imagem informal, é verdade, e confere status ontologicamente básico às propriedades mentais. Você não pode calcular quando ocorre uma “observação experimental” ou o que as pessoas “sabem”; você apenas sabe quando certas probabilidades são claramente zero. E esse “saber” se encaixa perfeitamente com os resultados experimentais, seja qual for o caso.

Mas, pelo menos, a consciência-causa-colapso pretende nos dizer como o universo funciona. As amplitudes são reais, o colapso é real, a consciência é real.

Comparado a este argumento:

ALUNO: “Espere, você está dizendo que esta amplitude desaparece quando uma medição indica que não é verdade?”

GOOFUS: “Não, não! Ela não desaparece literalmente. As equações não têm significado - elas simplesmente fazem boas previsões.”

ALUNO: “Mas então, o que acontece?”

GOOFUS: (irritado) “Nunca faça essa pergunta.”

ALUNO: “E o caso em que medimos a polarização deste fóton aqui, e, a um ano-luz de distância, a probabilidade de o fóton emaranhado estar polarizado de cima para baixo muda de 50% para 25%?”

GOOFUS: “Sim, e o que isso tem a ver?”

ALUNO: “Isso não viola a Relatividade Especial?”

GOOFUS: “Não, porque você está apenas determinando a polarização do outro fóton. Lembre-se, as amplitudes não são reais.”

ALUNO: “Mas o Teorema de Bell demonstra que não pode haver variáveis ocultas locais que descrevam a polarização do outro fóton antes da medição...”

GOOFUS: “Exatamente! Não faz sentido falar da polarização do fóton antes da medição.”

ALUNO: “Mas a probabilidade muda abruptamente...”

GOOFUS: “Não faz sentido discutir isso antes da medição!”

O que Goofus quer dizer com isso? Não importa o quão plausível possam parecer suas palavras; que tipo de realidade corresponderia a elas sendo verdadeiras?

De que forma a realidade deve ser para tornar sem sentido discutir a violação da Relatividade Especial porque a propriedade em questão não existe, mesmo que seja possível calcular mudanças nela?

Mas, sabe de uma coisa? Esqueça isso. Quero uma resposta para uma pergunta ainda mais crucial:

Onde Goofus obtém todas essas ideias?

Suponhamos que você pegue a equação de Schrödinger e a afirme como um fato positivo:

Esta equação produz boas previsões, mas não tem significado!

É mesmo? Como você sabe?

Às vezes, saio por aí afirmando que a questão fundamental da racionalidade é: por que você acredita no que acredita?

Você afirma que a equação de Schrödinger "não tem significado". Como você chegou a essa convicção, se não é simplesmente uma manifestação de ignorância interpretada erroneamente como conhecimento?

Houve algum experimento que lhe deu essa resposta? Estou disposto a considerar a possibilidade de que experimentos possam nos mostrar coisas que pareçam filosoficamente impossíveis. Contudo, neste caso, gostaria de ver os dados conclusivos. Houve um momento em que você montou cuidadosamente um aparato experimental e descobriu o que deveria esperar ver para decidir entre (1) a equação de Schrödinger ser significativa ou (2) a equação de Schrödinger ser desprovida de significado; e então você obteve o resultado (2)?

GALLANT: "Se eu medir a polarização de 90 graus de um fóton e, em seguida, medir a polarização de 45 graus e, em seguida, medir novamente a 90 graus, minha sequência de medições mostra que, em 100 tentativas, o fóton foi absorvido 47 vezes e transmitido 53 vezes."

GOOFUS: "A polarização de 90 graus e a polarização de 45 graus são propriedades incompatíveis; ambas não podem coexistir e, se você medir uma, não faz sentido falar sobre a outra."

Como você sabe?

De onde veio esse conhecimento, Goofus? Entendo de onde Gallant tirou o dele, mas e você?

Minha abordagem em relação a questões de existência e significado foi bem ilustrada em uma discussão sobre o estado atual das evidências sobre se o universo é espacialmente finito ou infinito, na qual James D. Miller repreendeu Robin Hanson:

Robin, você está exibindo excesso de confiança ao assumir que o universo existe. Certamente, há alguma possibilidade de o universo ter tamanho zero.

Minha resposta foi:

James, mesmo se o universo não existisse, ainda seria importante saber se ele é infinito ou finito.

Ah, você acha que usar o truque antigo de "o universo não existe" vai me impedir? Não vai nem me atrasar!

Não é que eu esteja descartando a possibilidade de o universo não existir. É que, mesmo na ausência de realidade, eu ainda desejo compreender o nada da melhor maneira possível. Minha curiosidade não desaparece de repente só porque a realidade está ausente, você sabe!

A natureza da "realidade" é algo que ainda me confunde, deixando a possibilidade de que ela não exista em aberto. Mas a Lei de Egan ainda se aplica: "Tudo se soma à normalidade". As maçãs continuam caindo, mesmo quando Einstein desafiou a teoria da gravidade de Newton.

Claro, pode acontecer que as maçãs não existam, a Terra não exista, a realidade não exista. No entanto, as maçãs inexistentes continuarão caindo em direção a um solo inexistente a uma taxa constante de 9,8 m/s^2.

Você afirma que o universo não existe? Muito bem, vou supor que eu acredite nisso - embora não esteja claro no que devo acreditar, além de repetir as palavras.

Agora, o que acontece se eu pressionar este botão?

Em "A Verdade Simples", eu disse:

Sinceramente, eu mesmo não tenho certeza de onde vem essa noção de "realidade". Não posso criar minha própria realidade no laboratório, então ainda não a compreendo completamente. No entanto, há momentos em que acredito fortemente que algo acontecerá e, em vez disso, ocorre outra coisa… Portanto, preciso de termos distintos para o que determina minhas previsões e o que determina meus resultados experimentais. Chamo as primeiras coisas de "crenças" e as últimas de "realidade".

Você quer dizer que as equações da mecânica quântica “não são reais”? Serei caridoso e vou supor que isso signifique algo. Mas o que isso poderia significar?

Talvez signifique que as equações que orientam minhas previsões sejam substancialmente diferentes daquelas que governam meus resultados experimentais. Então, o que determina meus resultados experimentais? Se você disser ser “nada”, quero entender que tipo de “nada” é esse e por que esse “nada” exibe uma notável regularidade ao determinar, por exemplo, minhas medições experimentais da massa de um elétron.

Não gosto muito de pessoas que me dizem para parar de fazer perguntas. Se você afirmar que algo definitivamente não faz sentido, quero saber exatamente o que você quer dizer com isso e como chegou a essa conclusão. Caso contrário, você não respondeu à minha pergunta; apenas me disse para parar de fazê-la.

“A Verdade Simples” narra a história de um pastor e seu aprendiz que descobrem como contar ovelhas jogando pedrinhas em baldes, quando são visitados por um delegado da corte curioso sobre o funcionamento dessas “pedrinhas mágicas”. O pastor tenta explicar: “Um balde vazio é mágico somente se os pastos estiverem desprovidos de ovelhas”, mas logo se vê envolvido nas discussões animadas do aprendiz e do delegado acerca de como a magia pode residir nas pedras.

Aqui, nos deparamos com equações quânticas que fornecem previsões experimentais excepcionais. O que exatamente significa para elas serem “desprovidas de sentido”? Seria como um balde de pedrinhas usado para contar ovelhas, mas que carece de magia?

Antes de o Teorema de Bell descartar as variáveis ocultas locais, era plausível, como Einstein acreditava, que pudesse existir uma descrição mais abrangente da realidade que ainda não possuíamos, e a teoria quântica era uma expressão do nosso conhecimento incompleto dessa descrição abrangente. As leis que aprendemos seriam semelhantes às leis da mecânica estatística: afirmações quantitativas da incerteza. Isso dificilmente tornaria as equações “desprovidas de sentido”; o conhecimento parcial é o significado da probabilidade.

Entretanto, o Teorema de Bell torna muito menos plausível que as equações quânticas representem um conhecimento parcial de algo determinístico, da mesma maneira que a mecânica estatística se relaciona com a física clássica como um conhecimento parcial de algo determinístico. E, mesmo assim, as equações quânticas não seriam “desprovidas de sentido” como essa expressão frequentemente sugere; elas seriam “estatísticas”, “aproximadas”, “informações parciais” ou, na pior das hipóteses, “incorretas”.

Aqui temos equações que nos proporcionam previsões excepcionais. Você alega que elas são “desprovidas de sentido”. Então, pergunto o que determina os resultados dos meus experimentos. E você não consegue responder. Bem, como, então, você justifica a eliminação da possibilidade de que as equações quânticas produzem previsões notáveis porque, digamos, são significativas?

Não cogito trivializar questões de realidade ou significado. Mas para rotular algo como “desprovido de sentido” e afirmar que a discussão está resolvida, encerrada, você deve possuir uma teoria precisa de como exatamente isso é desprovido de sentido. E, quando essa resposta for fornecida, a questão não deve mais parecer misteriosa.

Como você deve se lembrar das “Placas de Sinalização Semântica”, existem palavras e frases que não são respostas a perguntas, mas sim indicações de que você deve parar de fazer perguntas. “Por que algo existe em primeiro lugar? Deus!” é o exemplo clássico, mas há outros, como “elã vital!”

Dizer às pessoas para “calar a boca e calcular” porque você não sabe o que os cálculos significam e, em cinco anos, “calar a boca!” terá se disfarçado em uma teoria positiva da mecânica quântica.

Tenho grande respeito por qualquer físico histórico que tenha feito o esforço genuíno de realmente “calar a boca e calcular”, que tenha avaliado com sinceridade o que sabia e o que não sabia. Isso é o melhor que eles poderiam fazer sem serem realmente Hugh Everett, e eu lhes dou cinquenta pontos de racionalidade. Meu desprezo é reservado àqueles que interpretaram “Não sabemos por que funciona” como conhecimento positivo de que as equações definitivamente não eram reais.

Quero dizer, se esse truque funcionasse, seria bom demais para se limitar a um subcampo. Por que os físicos não deveriam usar a mesma tática fora da mecânica quântica?

"Ei, sua nova 'teoria do fio' não viola a Relatividade Especial?"

"Não, porque as equações não fazem sentido. E quanto ao seu modelo de 'inflação caótica do mal'? Não viola a simetria CPT?"

"Minhas equações são ainda mais sem sentido do que as suas!" Portanto, suas críticas em dobro não têm valor.

E se isso não funcionar, tente se dar um passe livre.

Se há uma lição a ser aprendida em toda essa história, é a lição de quão difícil é manter um estado de confusão reconhecida, sem criar uma narrativa que encerre o debate - o quanto é difícil evitar transformar a própria ignorância em conhecimento definido que se possui.

239 - Se muitos mundos tivesse vindo primeiro

Não que eu esteja afirmando que poderia ter feito melhor, se tivesse nascido naquela época, em vez desta…

A decoerência macroscópica, também conhecida como muitos mundos, foi proposta pela primeira vez em um artigo de 1957 por Hugh Everett III. O artigo foi ignorado. John Wheeler disse a Everett para ver Niels Bohr. Bohr não o levou a sério.

Arrasado, Everett deixou a física acadêmica, inventou o uso geral dos multiplicadores de Lagrange em problemas de otimização e se tornou multimilionário.

Só em 1970, quando Bryce DeWitt (que cunhou o termo "muitos mundos") escreveu um artigo para a Physics Today, que o campo geral foi informado pela primeira vez sobre as ideias de Everett. A decoerência macroscópica vem ganhando adeptos desde então e pode agora ser o ponto de vista majoritário (ou não).

Mas suponha que a decoerência e a decoerência macroscópica tivessem sido percebidas imediatamente após a descoberta do emaranhamento, na década de 1920. E suponha que ninguém tivesse proposto teorias de colapso até 1957. A decoerência estaria agora em constante declínio de popularidade, enquanto as teorias de colapso ganhariam força lentamente?

Imagine uma Terra alternativa, onde o primeiro físico a descobrir o emaranhamento e a superposição disse: "Santos macacos voadores, há um zilhão de outras Terras lá fora!"

Nos anos desde então, muitas hipóteses foram propostas para explicar as misteriosas probabilidades de Born. Mas ninguém ainda sugeriu um postulado de colapso. Essa possibilidade simplesmente não ocorreu a ninguém.

Um dia, Huve Erett entra no escritório de Biels Nohr…

"Eu simplesmente não entendo", disse Huve Erett, "por que ninguém na física parece sequer interessado na minha hipótese. As estatísticas de Born não são o maior enigma da teoria quântica moderna?"

Biels Nohr suspirou. Normalmente, ele nem se importaria, mas algo sobre o jovem o compeliu a tentar.

"Huve", diz Nohr, "todo físico encontra dezenas de pessoas por ano que acham que explicaram as estatísticas de Born. Se você vai a uma festa e diz a alguém que é físico, as chances são de pelo menos uma em dez de que eles tenham uma nova explicação para as estatísticas de Born. É um dos problemas mais famosos da ciência moderna e, pior, é um problema que todos acham que podem entender. Para chamar a atenção, uma nova hipótese de Born tem que ser…muito boa de verdade."

"E isso", diz Huve, "isso não é bom?"

Huve gesticula para o papel que trouxera para Biels Nohr. É um artigo curto. O título diz: "A Solução para o Problema de Born". O corpo do artigo diz:

Quando você realiza uma medição em um sistema quântico, todas as partes da função de onda,

exceto um ponto, desaparecem, com o sobrevivente escolhido não deterministicamente de uma maneira determinada pelas estatísticas de Born.

"Deixe-me ter certeza", diz Nohr cuidadosamente, "de que entendi você. Você está dizendo que temos essa função de onda - evoluindo conforme a equação de Wheeler-DeWitt - e, de repente, toda a função de onda, exceto uma parte, simplesmente vai para amplitude zero. Em todos os lugares ao mesmo tempo. Isso acontece quando, lá em cima, no nível macroscópico, nós 'medimos' algo."

"Certo!" diz Huve.

"Então a função de onda sabe quando nós a 'medimos'. O que exatamente é uma 'medição'? Como a função de onda sabe que estamos aqui? O que acontecia antes de os humanos estarem por perto para medir as coisas?"

"Hum…" Huve pensa por um momento. Então ele estende a mão para o papel, risca "Quando você realiza uma medição em um sistema quântico" e escreve: "Quando uma superposição quântica fica grande demais".

Huve olha para cima brilhantemente. "Consertado!"

"Entendo", diz Nohr. "E quão grande é 'grande demais'?"

"No nível de 50 mícrons, talvez", diz Huve, "Ouvi dizer que ainda não testaram isso."

De repente, um estudante coloca a cabeça na sala. "Ei, você ouviu? Eles acabaram de verificar a superposição no nível de 50 mícrons."

"Oh", diz Huve, "um, independentemente do nível, então. O que quer que faça os resultados experimentais saírem certos."

Nohr faz uma careta. "Olhe, rapaz, a verdade aqui não vai ser confortável. Você pode me ouvir sobre isso?"

"Sim", diz Huve, "eu só quero saber por que os físicos não me escutam."

"Tudo bem", diz Nohr. Ele suspira. "Olhe, se essa sua teoria fosse realmente verdadeira - se seções inteiras da função de onda simplesmente desaparecessem instantaneamente - seria… vejamos. A única lei em toda a mecânica quântica que é não-linear, não-unitária, não-diferenciável e descontínua. Impediria a física de evoluir localmente, com cada peça olhando apenas para seus vizinhos imediatos. Seu 'colapso' seria o único fenômeno fundamental em toda a física com uma base preferida e um espaço preferido de simultaneidade. O colapso seria o único fenômeno em toda a física que viola a simetria CPT, o Teorema de Liouville e a Relatividade Especial. Na sua versão original, o colapso também teria sido o único fenômeno em toda a física que era inerentemente mental. Deixei algo de fora?"

"O colapso também é o único fenômeno acausal", Huve aponta. "Isso não torna a teoria mais maravilhosa e incrível?"

"Eu acho, Huve", diz Nohr, "que os físicos podem ver o excepcionalismo da sua teoria como um ponto não a seu favor."

"Oh", disse Huve, surpreso. "Bem, acho que posso consertar essa coisa da não-diferenciabilidade postulando um termo de segunda ordem no-"

"Huve", diz Nohr, "acho que você não está entendendo meu ponto aqui. A razão pela qual os físicos não estão prestando atenção em você é que sua teoria não é física. É magia."

"Mas as estatísticas de Born são o maior enigma da física moderna, e esta teoria fornece um mecanismo para as estatísticas de Born!" Huve protesta.

“Não, Huve, não fornece”, diz Nohr cansado. “Isso é como dizer que você ‘forneceu um mecanismo’ para o eletromagnetismo dizendo haver pequenos anjos empurrando as partículas carregadas conforme as equações de Maxwell. Em vez de dizer: “Aqui estão as equações de Maxwell, que dizem aos anjos onde empurrar os elétrons”, nós simplesmente dizemos: ‘Aqui estão as equações de Maxwell’ e ficamos com uma teoria estritamente mais simples. Agora, não sabemos por que as estatísticas de Born acontecem. Mas você não deu a menor razão para que seu ‘postulado de colapso’ elimine mundos conforme as estatísticas de Born, em vez de outra coisa. Você nem está usando o fato de que a evolução quântica é unitária-”

“Isso é porque não é”, interrompe Huve.

“...o que todo mundo praticamente sabe que tem que ser a chave para as estatísticas de Born, de alguma forma. Em vez disso, você está meramente dizendo: ‘Aqui estão as estatísticas de Born, que dizem ao colapsador como eliminar mundos’, e é estritamente mais simples apenas dizer ‘Aqui estão as estatísticas de Born’.”

“Mas…” diz Huve.

“Além disso”, diz Nohr, elevando a voz, “você não deu nenhuma justificativa para por que há apenas um mundo sobrevivente deixado pelo colapso, ou por que o colapso acontece antes que qualquer humano fique superposto, tornando sua teoria realmente suspeita para um físico moderno. Este é exatamente o tipo de hipótese não testável que a multidão ‘Um Cristo’ usa para argumentar que devemos ‘ensinar a controvérsia’ quando falamos aos alunos do ensino médio sobre outras Terras.”

“Eu não sou um… Um-Cristista!” protesta Huve.

“Ótimo”, diz Nohr, “então por que você simplesmente assume que só resta um mundo? E esse não é o único problema com sua teoria. Que parte da função de onda é eliminada, exatamente? E em qual base? É claro que a função de onda inteira não está sendo comprimida para um delta, ou computadores quânticos comuns não poderiam permanecer em superposição quando qualquer colapso ocorresse em qualquer lugar - caramba, a química molecular comum poderia começar a falhar-”

Huve rapidamente risca “um ponto” em seu papel, escreve “uma parte” e então diz: “O colapso não comprime a função de onda para um ponto. Ele elimina toda a amplitude exceto um mundo, mas deixa toda a amplitude nesse mundo.”

“Por quê?” diz Nohr. “Em princípio, uma vez que você postula ‘colapso’, então ‘colapso’ poderia eliminar qualquer parte da função de onda, em qualquer lugar - por que apenas um mundo arrumado deixado? O colapsador sabe que estamos aqui dentro?”

Huve diz: “Ele deixa um mundo inteiro porque é isso que se encaixa em nossos experimentos.”

“Huve”, diz Nohr pacientemente, “o termo para isso é ‘post hoc’. Além disso, a decoerência é um processo contínuo. Se você particionar por cérebros inteiros com neurônios distintos disparando, as partições têm quase zero interferência mútua na função de onda. Mas muitos outros processos se sobrepõem muito. Não há nenhuma maneira possível de você apontar para ‘um mundo’ e eliminar o resto sem fazer escolhas completamente arbitrárias, incluindo uma escolha arbitrária de base-”

“Mas-” diz Huve.

“E acima de tudo”, diz Nohr, “a razão pela qual você não pode me dizer, qual parte da função de onda desaparece, ou exatamente quando isso acontece, ou exatamente o que a desencadeia, é que se adotássemos essa sua teoria, seria a única lei fundamental informalmente especificada e qualitativa ensinada em toda a física. Logo, dois físicos em qualquer lugar concordariam sobre os detalhes exatos! Por quê? Porque seria a única lei fundamental em toda a física moderna que seria acreditada sem evidência experimental para determinar exatamente como funciona.”

“O quê, sério?” diz Huve. “Pensei que muita física fosse mais informal que isso. Quer dizer, você não estava falando sobre como é impossível apontar para ‘um mundo’?”

“Isso é porque mundos não são fundamentais, Huve! Temos evidências experimentais massivas sustentando a lei fundamental, a equação de Wheeler-DeWitt, que usamos para descrever a evolução da função de onda. Nós apenas aplicamos a mesma equação para obter nossa descrição da decoerência macroscópica. Não fosse pelas dificuldades de cálculo, a equação nos diria, em princípio, exatamente quando ocorreu a decoerência macroscópica. Não sabemos de onde vêm as estatísticas de Born, mas temos evidências massivas do que são as estatísticas de Born. Mas quando pergunto a você quando, ou onde, ocorre o colapso, você não sabe - porque não há absolutamente nenhuma evidência experimental para determinar isso. Huve, mesmo se esse ‘postulado de colapso’ funcionasse da maneira que você diz que funciona, não há nenhuma maneira possível de você saber disso! Por que não um zilhão de outras possibilidades igualmente mágicas?”

Huve levanta as mãos defensivamente. “Não estou dizendo que minha teoria deveria ser ensinada nas universidades como verdade aceita! Eu só quero que ela seja testada experimentalmente! Isso é tão errado?”

“Você não especificou quando o colapso acontece, então não posso construir um teste que falsifique sua teoria”, diz Nohr. “Agora, dito isso, já estamos procurando experimentalmente por qualquer parte das leis quânticas que mude em níveis cada vez mais macroscópicos. Tanto por princípios gerais, caso haja algo na 20ª casa decimal que só apareça em sistemas macroscópicos, quanto na esperança de descobrirmos algo que lance luz sobre as estatísticas de Born. Verificamos os tempos de decoerência como uma questão de rotina. Mas mantemos uma visão ampla sobre o que pode ser diferente. Ninguém vai privilegiar seu ‘colapso’ não-linear, não-unitário, não-diferenciável, não-local, não-simétrico-CPT, não-relativístico, a-causal, mais rápido que a luz e informal quando se trata de procurar pistas. Não até que vejam evidências absolutamente inconfundíveis. E acredite em mim, Huve, será preciso um inferno de evidências para não confundir isso. Mesmo se encontrássemos tempos de decoerência anômalos, e não acho que encontraremos, isso não forçaria seu ‘colapso’ como explicação.”

“O quê?” diz Huve. “Por que não?”

“Porque deve haver um bilhão de explicações mais plausíveis do que violar a Relatividade Especial”, diz Nohr. “Você percebe que se isso realmente acontecesse, haveria apenas um único resultado quando você medisse a polarização de um fóton? Medir um fóton em um par emaranhado influenciaria o outro fóton a um ano-luz de distância. Einstein teria um ataque cardíaco.”

“Não viola realmente a Relatividade Especial”, diz Huve. “O colapso ocorre exatamente da maneira certa para impedir que você detecte a influência mais rápida que a luz.”

“Isso não é um ponto a favor da sua teoria”, diz Nohr. “Além disso, Einstein ainda teria um ataque cardíaco.”

“Oh”, diz Huve. “Bem, diremos que os aspectos relevantes da partícula não existem até que o colapso ocorra. Se algo não existe, influenciá-lo não viola a Relatividade Especial-”

“Você só está se enterrando mais fundo. Olhe, Huve, como princípio geral, teorias que são realmente corretas não geram esse nível de confusão. Mas acima de tudo, não há nenhuma evidência para isso. Você não tem nenhuma maneira lógica de saber que o colapso ocorre, e nenhuma razão para acreditar nisso. Você cometeu um erro. Apenas diga ‘opa’ e siga com sua vida.”

“Mas eles poderiam encontrar a evidência algum dia”, diz Huve.

“Não consigo imaginar que evidência poderia determinar essa hipótese particular de um-mundo como explicação, mas de qualquer forma, agora não encontramos nenhuma evidência dessas”, diz Nohr. “Não encontramos nada nem vagamente sugestivo disso! Você não pode atualizar com base em evidências que poderiam teoricamente chegar algum dia, mas não chegaram! Agora, hoje, não há razão para gastar tempo valioso pensando nisso em vez de um bilhão de outras teorias igualmente mágicas. Não há absolutamente nada que justifique sua crença na ‘teoria do colapso’ mais do que acreditar que algum dia aprenderemos a transmitir mensagens mais rápidas que a luz aproveitando os efeitos não causais de rezar para o Monstro do Espaguete Voador!”

Huve se endireita com dignidade ferida. “Você sabe, se minha teoria estiver errada - e admito que pode estar errada-”

“Se?” diz Nohr. “Pode?”

“Se, eu digo, minha teoria estiver errada”, Huve continua, “então em algum lugar lá fora há outro mundo onde eu sou o físico famoso e você é o único pária!”

Nohr enterra a cabeça nas mãos. “Oh, não isso de novo. Você não ouviu o ditado ‘Viva em seu próprio mundo’? E você entre todas as pessoas-”

“Em algum lugar lá fora há um mundo onde a grande maioria dos físicos acredita na teoria do colapso, e ninguém sequer sugeriu decoerência macroscópica nos últimos trinta anos!”

Nohr levanta a cabeça e começa a rir.

“O que é tão engraçado?” Huve diz desconfiado.

Nohr só ri mais forte. “Oh, meu Deus! Oh, meu Deus! Você realmente acha, Huve, que há um mundo lá fora onde eles conhecem a física quântica há trinta anos, e ninguém nem pensou que poderia haver mais de um mundo?”

“Sim”, diz Huve, “é exatamente isso que eu penso.”

“Oh, meu Deus! Então você está dizendo, Huve, que os físicos detectam superposição em sistemas microscópicos e elaboram equações quantitativas que governam a superposição em cada instância que podem testar. E por trinta anos, nem uma pessoa diz: ‘Ei, eu me pergunto se essas leis por acaso são universais’.”

“Por que deveriam?” diz Huve. “Modelos físicos às vezes se mostram errados quando você examina novos regimes.”

“Mas nem sequer pensar nisso?” Nohr diz incredulamente. “Você vê maçãs caindo, elabora a lei da gravidade para todos os planetas do sistema solar, exceto Júpiter, e nem lhe ocorre aplicá-la a Júpiter porque Júpiter é grande demais? Isso é como, como uma espécie de esquete de comédia onde o cara abre uma caixa, e ela contém uma torta com mola, então o cara abre outra caixa, e ela contém outra torta com mola, e o cara simplesmente continua fazendo isso sem nem pensar na possibilidade de que a próxima caixa também contenha uma torta. Você acha que John von Neumann, que pode ter sido o humano com o QI mais alto da história, não pensaria nisso?”

“Isso mesmo”, diz Huve, “Ele não pensaria. Pondere sobre isso.”

“Este é o mundo onde meu bom amigo Ernest formula seu experimento mental do Gato de Schrödinger, e neste mundo, o experimento mental é assim: ‘Ei, suponha que temos uma partícula radioativa que entra em uma superposição de decair e não decair. Então a partícula interage com um sensor, e o sensor entra em uma superposição de disparar e não disparar. O sensor interage com um explosivo, que entra em uma superposição de explodir e não explodir; que interage com o gato, então o gato entra em uma superposição de estar vivo e morto. Então um humano olha para o gato,’ e neste ponto Schrödinger para e diz: ‘poxa, eu simplesmente não consigo imaginar o que poderia acontecer a seguir.’ Então Schrödinger mostra isso para todos os outros, e eles também ficam, tipo ‘Uau, não faço ideia do que poderia acontecer neste ponto, que paradoxo incrível.’ Até que finalmente você ouve sobre isso, e você fica tipo, ‘ei, talvez nesse ponto metade da superposição simplesmente desapareça, aleatoriamente, mais rápido que a luz,’ e todos os outros ficam, tipo, ‘Uau, que ideia incrível!’ “

“Isso mesmo”, diz Huve novamente. “Tem que ter acontecido em algum lugar.”

“Huve, este é um mundo onde cada físico, e provavelmente toda a maldita espécie humana, é burra demais para se inscrever na criônica! Estamos falando da Terra onde George W. Bush é Presidente.”

240 - Onde a filosofia encontra a ciência

Olhando para trás, para os primórdios da física quântica - não com o propósito de criticar as principais figuras, ou afirmar que poderíamos ter feito melhor se tivéssemos nascido naquela era, mas para tentar aprender uma lição moral e fazer melhor da próxima vez - olhando para as eras sombrias da física quântica, digo, eu nomearia como o "erro mais básico"…

… Não o fato de que tentaram reverter os últimos três mil anos de ciência sugerindo que a mente era complexa na física, em vez de fundamental na física. Isso é Ciência, e temos revoluções aqui. Ocasionalmente você precisa reverter uma tendência. O futuro é sempre absurdo e nunca ilegal.

Eu nomearia, como o erro básico a não repetir da próxima vez, que os primeiros cientistas esqueceram que eles próprios eram feitos de partículas.

Quero dizer, tenho certeza de que a maioria deles sabia disso em teoria.

E ainda assim, eles não perceberam que colocar um sensor para detectar um elétron passando, ou mesmo saber sobre a história do elétron, era um exemplo de "partículas em lugares diferentes". Então, eles não notaram que uma teoria quântica de configurações distintas já explicava o resultado experimental, sem necessidade de invocar a consciência.

No ambiente ancestral, os humanos enfrentavam frequentemente a tarefa adaptativamente relevante de prever outros humanos. Para isso, você pensava em seus companheiros humanos como tendo pensamentos, sabendo coisas e sentindo coisas, em vez de pensar neles como feitos de partículas. De fato, muitas tribos de caçadores-coletores podem nem ter sabido que partículas existiam. É muito mais intuitivo—parece mais simples—pensar em alguém "sabendo" algo, do que pensar nas partículas do cérebro deles ocupando um estado diferente. É mais fácil formular suas explicações em termos do que as pessoas sabem; parece mais natural; vem mais prontamente à mente.

Assim como, antigamente, era mais fácil imaginar Thor lançando raios, do que imaginar as Equações de Maxwell—mesmo que as Equações de Maxwell possam ser descritas por um programa vastamente menor que o programa para um agente inteligente como Thor.

Então, os antigos físicos acharam natural pensar: "Eu sei onde o fóton estava… que diferença isso poderia fazer?" Não: "O estado atual das partículas do meu cérebro correlaciona-se com a história do fóton… que diferença isso poderia fazer?"

E, da mesma forma, porque parecia fácil e intuitivo modelar a realidade em termos do que as pessoas sabem, e a decomposição do saber em estados cerebrais não surgia tão prontamente, parecia uma teoria simples dizer que uma configuração poderia ter amplitude apenas "se você não soubesse melhor".

Para transformar a hipótese quântica dualista em uma teoria formal—uma que pudesse ser escrita como um programa, sem cientistas humanos decidindo quando ocorreu uma "observação"—você teria que especificar o que significava para um "observador" "saber" algo, em termos que seu programa pudesse computar.

Então, sua teoria da física fundamental examinará todas as partículas em um cérebro humano, e

decidir quando essas partículas “sabem” algo, para calcular os movimentos das partículas? Mas como você calcula o movimento das partículas no próprio cérebro? Não haveria uma potencial recursão infinita?

Mas enquanto os termos da teoria estavam sendo processados por cientistas humanos, eles simplesmente sabiam quando ocorreu uma “observação”. Você dizia que uma “observação” ocorreu sempre que precisava ocorrer para as previsões experimentais saírem certas—uma forma sutil de ajuste constante.

(Lembre-se, os fundamentos da teoria quântica foram formulados antes de Alan Turing dizer qualquer coisa sobre máquinas de Turing, e muito antes do conceito de computação ser popularmente conhecido. A distinção entre uma teoria formal eficaz, e uma que requeria interpretação humana, não era tão clara então quanto é agora. Fácil identificar os problemas em retrospectiva; você não deve aprender a lição de que os problemas geralmente são tão óbvios de antemão.)

Olhando para trás, pode parecer que uma meta-lição a aprender com a história é que a filosofia realmente importa na ciência—não é apenas algum complemento de um campo acadêmico separado.

Afinal, os primeiros cientistas quânticos estavam fazendo todos os experimentos corretos. Foi a interpretação deles que estava errada. E os problemas de interpretação não foram o resultado de eles errarem as estatísticas.

Olhando para trás, parece que os erros que cometeram foram erros no tipo de pensamento que descrevemos como, bem, “filosófico”.

Quando olhamos para trás e perguntamos: “Como os primeiros cientistas quânticos poderiam ter feito melhor, mesmo em princípio?” parece que os insights que precisavam eram filosóficos.

E ainda assim, não foram filósofos profissionais que resolveram o problema e esclareceram o mistério e tornaram tudo normal novamente. Foram, bem, físicos.

Discutivelmente, Leibniz foi pelo menos tão visionário sobre a física quântica quanto Demócrito foi uma vez considerado visionário sobre átomos. Mas isso é retrospectiva. É o resultado de olhar a solução e pensar para trás, e dizer: “Ei, Leibniz disse algo assim.”

Mesmo onde um filósofo acerta com antecedência, é geralmente a ciência que acaba nos dizendo qual filósofo estava certo—não o consenso anterior da comunidade filosófica.

Acho que isso tem algo fundamental a dizer sobre a natureza da filosofia e a interface entre filosofia e ciência.

Já foi dito que toda ciência começa como filosofia, mas depois cresce e deixa o útero filosófico, de modo que, a qualquer momento, “Filosofia” é o que ainda não transformamos em ciência.

Sugiro que, quando olhamos para a história da física quântica e dizemos: “Os insights de que precisavam eram insights filosóficos”, o que realmente estamos vendo é que o insight de que precisavam era de uma forma que ainda não é ensinada em aulas acadêmicas padronizadas, e ainda não foi reduzida a cálculos.

Era uma vez, a noção de método científico - atualizar crenças com base em evidências experimentais - era uma noção filosófica. Mas não foi defendida por filósofos profissionais. Foi o poder da realidade da ciência que mostrou que a epistemologia científica era uma boa epistemologia, não um consenso anterior de filósofos.

Hoje, essa filosofia de atualização de crenças está começando a ser reduzida a cálculos - estatística, teoria da probabilidade bayesiana.

Mas na era de Galileu, eram apenas argumentos verbais vagos que diziam que você deveria tentar produzir previsões numéricas de resultados experimentais, em vez de consultar a Bíblia ou Aristóteles.

Na fronteira da ciência, e especialmente na fronteira do caos científico e da confusão científica, você encontra problemas de pensamento que não são ensinados em cursos acadêmicos, e que não foram re-

duzidos a cálculos. E isso parecerá um domínio da filosofia; parecerá que você deve fazer um pensamento filosófico para resolver a confusão. Mas quando a história olha para trás, temo, geralmente não é um filósofo profissional que ganha todas as fichas - porque é necessário um envolvimento íntimo com o domínio científico para fazer o pensamento filosófico. Mesmo que, depois, tudo pareça conhecível a priori; e mesmo que, depois, algum filósofo por aí tenha realmente acertado a priori; mesmo assim, é necessário um envolvimento íntimo para ver isso, na prática, e resultados experimentais para dizer ao mundo qual filósofo venceu.

Sugiro que, como a ética, a filosofia realmente é importante, mas é praticada eficazmente apenas numa ciência. Tentar fazer a filosofia de uma ciência de fronteira, como uma profissão acadêmica separada, é um erro tão grande quanto tentar ter éticos separados. Você acaba com éticos que falam principalmente com outros éticos, e filósofos que falam principalmente com outros filósofos.

Isso não quer dizer que não haja lugar para filósofos profissionais no mundo. Alguns problemas são tão caóticos que não há lugar estabelecido para eles nos salões da ciência. Mas esses "filósofos profissionais" seriam muito, muito sábios se aprendessem cada fragmento de ciência relevante que pudessem conseguir. Eles não deveriam se surpreender com a perspectiva de que o experimento, e não o debate, resolverá finalmente a argumentação. Eles não deveriam hesitar em conduzir seus próprios experimentos, se puderem pensar em algum.

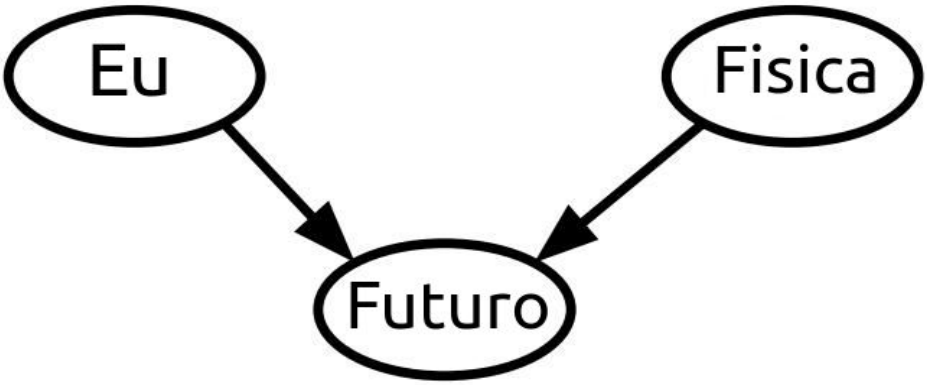
Essa, eu acho, é a lição de história.

241 - Tu és Física

Há três meses — caramba, já faz tanto tempo assim? — propus o seguinte dever de casa: Faça um rastreamento das pilhas de execução dos algoritmos cognitivos humanos que produzem debates sobre o "livre-arbítrio". Note que esta tarefa é fortemente diferenciada de argumentar se o livre-arbítrio existe ou não.

Agora, como esperado, as pessoas estão perguntando: "Se o futuro é determinado, como nossas escolhas podem controlá-lo?" O leitor sábio pode adivinhar que tudo se soma à normalidade; mas isso deixa a questão de como.

As pessoas ouvem: "O universo funciona como um relógio; a física é determinística; o futuro é fixo." E suas mentes formam uma rede causal que se parece com isto:



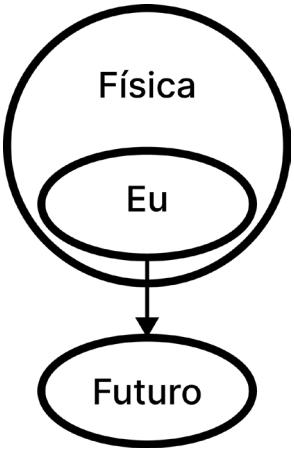
Aqui vemos as causas "Eu" e "Física" competindo para determinar o estado do efeito "Futuro". Se o "Futuro" é totalmente determinado pela "Física", então obviamente não há espaço para ser afetado por "Eu".

Esta rede causal não é uma crença filosófica explícita. É implícita - uma representação de fundo do cérebro, controlando quais argumentos filosóficos parecem "razoáveis". Simplesmente parece ser a maneira como as coisas são.

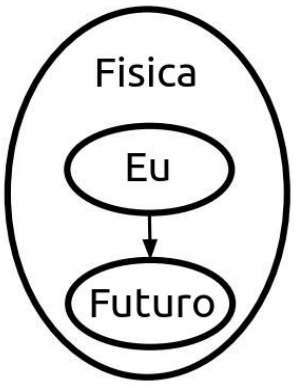
Ocasionalmente, surge outro comunicado de imprensa sobre neurociência, alegando que, porque os pesquisadores usaram uma ressonância magnética funcional para detectar o cérebro fazendo algo durante um processo de decisão, não é você quem escolhe, é seu cérebro.

Da mesma forma, aquela velha história: "O reducionismo mina a própria racionalidade. Porque então, cada vez que você dissesse algo, não seria o resultado de raciocinar sobre as evidências — seriam meramente quarks saltitando por aí."

Claro que o diagrama real deveria ser:



Ou melhor ainda:



Por que isso não é óbvio? Porque há muitos níveis de organização que separam nossos modelos de nossos pensamentos — nossas emoções, nossas crenças, nossas indecisões agonizantes e nossas escolhas finais — de nossos modelos de elétrons e quarks.

Podemos visualizar intuitivamente que uma mão é feita de dedos (e polegar e palma). Perguntar se é realmente nossa mão que pega algo, ou meramente nossos dedos, polegar e palma, é transparentemente uma pergunta errada.

Mas o abismo entre a física e a cognição não pode ser cruzado por visualização direta. Ninguém pode visualizar átomos formando uma pessoa, da maneira como podem ver dedos formando uma mão.

E assim, é necessária vigilância constante para manter sua percepção de si como uma entidade na física.

Esta vigilância é uma das grandes chaves para a filosofia, como a Falácia da Projeção Mental. Você se lembrará que foi este ponto que nomeei como tendo feito tropeçar os físicos quânticos que falharam em imaginar a decoerência macroscópica; eles não pensaram em aplicar as leis neles mesmos.

Crenças, desejos, emoções, valores morais, objetivos, imaginações, antecipações, percepções sensoriais, desejos fugazes, ideais, tentações… Você poderia chamar isso de “camada superficial” da mente, as partes do eu que as pessoas podem ver mesmo sem ciência. Se eu disser: “Não é você quem determina o futuro, são seus desejos, planos e ações que determinam o futuro”, você pode facilmente ver as relações parte-todo. É imediatamente visível, como dedos formando uma mão. Há outras relações parte-todo até chegar à física,

mas elas não são imediatamente visíveis.

O “Compatibilismo” é a posição filosófica de que o “livre-arbítrio” pode ser definido de forma intuitiva e satisfatória de modo a ser compatível com a física determinística. O “Incompatibilismo” é a posição de que o livre-arbítrio e o determinismo são incompatíveis.

Minha posição talvez possa ser chamada de “Requerimentismo”. Quando a agência, a escolha, o controle e a responsabilidade moral são compreendidos de maneira sensata, eles requerem determinismo — pelo menos alguns patches de determinismo no universo. Se você escolhe, planeja, age e traz algum futuro à existência, de acordo com seu desejo, então tudo isso requer um tipo de realidade regida por leis; você não pode fazer isso em meio ao caos total. Deve haver ordem pelo menos sobre aquelas partes da realidade que estão sendo controladas por você. Você está na física, e assim você/física determinaram o futuro. Se não fosse determinado pela física, não poderia ser determinado por você.

Ou talvez eu deva dizer: “Se o futuro não fosse determinado pela realidade, não poderia ser determinado por você”, ou “Se o futuro não fosse determinado por algo, não poderia ser determinado por você”. Você não precisa de neurociência ou física para empurrar definições ingênuas de livre-arbítrio para a incoerência. Se a mente não estivesse incorporada no cérebro, estaria incorporada em outra coisa; haveria alguma coisa real que seria uma mente. Se o futuro não fosse determinado pela física, seria determinado por algo, alguma lei, alguma ordem, alguma grande realidade que incluísse você dentro dela.

Mas se as leis da física nos controlam, então como podemos dizer que controlamos a nós mesmos?

Inverta a questão: Se as leis da física não nos controlassem, como poderíamos possivelmente controlar a nós mesmos?

Como os pensamentos poderiam julgar outros pensamentos, como as emoções poderiam entrar em conflito entre si, como um curso de ação poderia parecer melhor, como poderíamos passar da incerteza para a certeza sobre nossos próprios planos, em meio ao caos total?

Se não estivéssemos na realidade, onde poderíamos estar?

O futuro é determinado pela física. Que tipo de física? O tipo de física que inclui as ações dos seres humanos.

As escolhas das pessoas são determinadas pela física. Que tipo de física? O tipo de física que inclui pesar decisões, considerar possíveis resultados, julgá-los, ser tentado, seguir valores morais, racionalizar transgressões, tentar fazer melhor…

Não há um ponto onde um quark vem voando de Plutão e anula tudo isso.

Os pensamentos do seu processo de decisão são todos reais, todos são algo. Mas um pensamento é grande e complicado demais para ser um átomo. Então os pensamentos são feitos de coisas menores, e nosso nome para a matéria de que a matéria é feita é “física”.

A física fundamenta nossas decisões e inclui nossas decisões. Ela não as explica.

Lembre-se, a física se soma à normalidade; são seus algoritmos cognitivos que geram confusão.

242 - Muitos mundos, uma melhor suposição

Se você observar muitos fenômenos físicos microscópicos—um fóton, um elétron, um átomo de hidrogênio, um laser—e um milhão de outros experimentos conhecidos—é possível formular leis simples que parecem governar todas as pequenas coisas (contanto que você não pergunte sobre a gravidade). Essas leis governam a evolução de um objeto altamente abstrato e matemático que chamei de "distribuição de amplitude", mas que é mais amplamente conhecido como "função de onda".

Agora, existem questões complexas sobre a generalização apropriada que cobre todos esses casos minúsculos. Chame um objeto de "verzul" se ele parecer verde antes de 1º de janeiro de 2020 e parecer azul depois dessa data. Se todos os esmeraldas examinados até agora pareceram verdes, a generalização correta é "Esmeraldas são verdes" ou "Esmeraldas são verzuis"?

A resposta é que a generalização correta é "Esmeraldas são verdes". Não entrarei nos argumentos agora. Esse não é o assunto deste ensaio, e a resposta óbvia nesse caso acontece de estar correta. O verdadeiro Caminho não é estúpido: por mais inteligente que você seja com sua lógica, ela deve finalmente chegar à resposta certa e não à errada.

Da mesma forma, as generalizações mais simples que cobririam apenas os fenômenos microscópicos observados tomam a forma de "Todos os elétrons têm spin 1/2" e não "Todos os elétrons têm spin 1/2 antes de 1º de janeiro de 2020" ou "Todos os elétrons têm spin 1/2, a menos que façam parte de um sistema emaranhado que pese mais de 1 grama".

Quando voltamos nossa atenção para fenômenos macroscópicos, nossa visão fica obscurecida. Não podemos experimentar a função de onda de um humano da mesma forma que podemos experimentar a função de onda de um átomo de hidrogênio. Em nenhum caso você pode realmente ler a função de onda com um pequeno scanner quântico. Mas, no caso de, por exemplo, um humano, o tamanho do organismo inteiro impede nossa capacidade de realizar cálculos ou experimentos precisos—não podemos confirmar que as equações quânticas estão sendo obedecidas em detalhes precisos.

Sabemos que fenômenos comumente considerados "quânticos" não desaparecem apenas porque muitos objetos microscópicos são agregados. Lasers emitem um fluxo de fótons coerentes, em vez de, digamos, fazer algo completamente diferente. Os átomos têm as características químicas que a teoria quântica diz que devem ter, permitindo que se agreguem em moléculas estáveis que compõem um humano.

Então, em um sentido, temos inúmeras evidências de que as leis quânticas estão se agregando ao nível macroscópico sem muita diferença. A química em massa ainda funciona.

Mas não podemos verificar diretamente que as partículas que compõem um humano têm uma função de onda agregada que se comporta exatamente da maneira que as leis quânticas mais simples dizem. Ah, sabemos que moléculas e átomos não se desintegram, sabemos que espelhos macroscópicos ainda refletem do meio. Podemos obter muitas previsões de alto nível a partir da suposição de que o microscópico e o macroscópico são governados pelas mesmas leis, e toda previsão testada se provou verdadeira.

Mas se alguém alegar que a imagem quântica macroscópica difere da microscópica em algum detalhe ainda não testável—algo que só aparece na vigésima casa decimal das interações microscópicas, mas se agrega em algo maior para interações macroscópicas—bem, não podemos provar que estão errados. É a Navalha

de Ocam que diz: “Existem zilhões de novas leis fundamentais que você poderia postular na vigésima casa decimal; por que você está pensando nesta em particular?”

Se calcularmos usando as leis mais simples que governam todos os casos conhecidos, descobrimos que os humanos acabam em estados de superposição quântica, assim como fótons em uma superposição de refletir e passar através de um espelho semi-refletor. No experimento do Gato de Schrödinger, um átomo instável entra em uma superposição de se desintegrar e não se desintegrar. Um sensor, ajustado para o átomo, entra em uma superposição de disparar e não disparar. (Na verdade, a superposição é agora um estado conjunto de [átomo desintegrado + sensor disparado] + [átomo estável + sensor não disparado]). Uma carga de explosivos, conectada ao sensor, entra em uma superposição de explodir e não explodir; um gato na caixa entra em uma superposição de estar morto e vivo; e um humano, olhando na caixa, entra em uma superposição de vomitar e estar calmo. A mesma lei em todos os níveis.

Os seres humanos que interagem com sistemas em superposição irão, eles próprios, evoluir para superposições. Mas o cérebro que vê o gato explodido, e o cérebro que vê o gato vivo, terão muitos neurônios disparando de maneiras diferentes, e, portanto, muitas partículas em diferentes posições. Eles estão muito distantes no espaço de configuração e se comunicarão em um grau exponencialmente infinitesimal. Não na trigésima casa decimal, mas na casa 10^{30}. Nenhuma mente em particular, nenhum processo cognitivo em particular, vê uma superposição borrada de gatos.

O fato de que “você” só parece ver o gato vivo ou morto é exatamente o que as leis quânticas mais simples preveem. Portanto, não temos razão para acreditar, a partir de nossa experiência até agora, que as leis quânticas são de alguma forma diferentes no nível macroscópico do que no nível microscópico.

E os físicos têm verificado a superposição em níveis crescentes. Aparentemente, um esforço está atualmente em andamento para testar a superposição em um objeto de 50 mícrons, maior do que a maioria dos neurônios.

A existência de outras versões de nós mesmos, e de fato outras Terras, não é suposta adicionalmente. Estamos simplesmente supondo que as mesmas leis governem em todos os níveis, não tendo razão para supor o contrário, e todos os testes experimentais tendo sido bem-sucedidos até agora. A existência de outras Terras decoerentes é uma consequência lógica da generalização mais simples que se encaixa em todos os fatos conhecidos. Se você acha que a Navalha de Ocam diz que os outros mundos são “entidades desnecessárias” sendo multiplicadas, então você deve verificar a matemática da teoria das probabilidades; não é assim que a Navalha de Ocam funciona.

No entanto, há um enigma particular que parece estranho ao tentar estender as leis microscópicas universalmente, incluindo para humanos em superposição:

Se tentarmos obter probabilidades contando o número de observadores distintos, não há razão óbvia pela qual o módulo quadrado integrado da função de onda deve correlacionar-se com resultados experimentais estatísticos. Não há razão conhecida para as probabilidades de Born, e parece que, a priori, esperaríamos uma probabilidade de 50/50 de qualquer experimento quântico binário indo para ambos os lados, se só contássemos os observadores.

Robin Hanson sugere que se bolhas de amplitude decoerentes exponencialmente menores que a média (“mundos”) são interferidos por vazamentos exponencialmente pequenos de bolhas maiores, obteremos as probabilidades de Born de volta. Considero isso uma possibilidade interessante, porque é tão normal.

(Eu mesmo tive pensamentos recentes em uma direção diferente: se tento contar observadores da maneira óbvia, obtenho resultados que parecem estranhos em geral, não apenas no caso da física quântica. Se, por exemplo, eu dividir meu cérebro em um trilhão de partes semelhantes, condicionado a ganhar na loteria enquanto anestesiado; permitir que meus “eus” acordem e talvez se diferenciem em pequenos graus uns dos outros; e depois mesclá-los novamente em um único eu; então contar observadores da maneira óbvia diz que eu deveria conseguir me fazer ganhar na loteria (se eu puder dividir meu cérebro e mesclá-lo, como uma mente carregada poderia conseguir fazer).

Nesse contexto, acho muito interessante que a regra de Born não tenha um problema de dividir e remesclar. Dada a física quântica unitária, a regra de Born é a regra única que impede "observadores" de terem poderes psíquicos—o que não explica a regra de Born, mas certamente é um fato interessante. Dada a regra de Born, mesmo dividir e mesclar mundos ainda levaria a probabilidades consistentes. Talvez a física use uma IA melhor do que eu uso!

Talvez eu devesse seguir as dicas da física, em vez de tentar raciocinar a priori, e ver onde isso me leva? Mas não fui levado a lugar nenhum ainda, então isso dificilmente é uma "resposta".)

Wallace, Deutsch e outros tentam derivar a Regra de Born a partir da teoria da decisão. Sou bastante suspeito disso, porque parece que há um componente de "O que acontece comigo?" que não posso alterar modificando minha função utilidade. Mesmo se eu não me importasse nada com mundos onde não ganhei na loteria quântica, ainda parece haver um sentido em que eu "ma maioria das vezes" acordaria em um mundo onde não ganhei na loteria. É isso que acho que precisa ser explicado.

O ponto é que muitas hipóteses sobre as probabilidades de Born foram propostas. Não tantas quanto deveriam, porque o mistério foi falsamente marcado como "resolvido" por um longo tempo. Mas ainda assim, houveram muitas propostas.

Há uma esperança legítima de uma solução para o enigma de Born sem novas leis fundamentais. Seu mundo não se divide em exatamente dois novos subprocessos na ocasião exata em que você vê "absorvido" ou "transmitido" na tela LCD de um sensor de fóton. Estamos constantemente sendo superpostos e decoeridos, o tempo todo, às vezes ao longo de dimensões contínuas - embora os cérebros sejam digitais e envolvam neurônios inteiros disparando, e disparar/não-disparar seria um estado extremamente decoerente até mesmo de um único neurônio... Parece haver espaço para algo inesperado explicar as estatísticas de Born - uma melhor compreensão do peso antrópico dos observadores, ou uma melhor compreensão das superposições do cérebro - sem novos fundamentos.

Não podemos descartar, no entanto, a possibilidade de que uma nova lei fundamental esteja envolvida nas estatísticas de Born.

Como Jess Riedel coloca:

> Se há uma lição que podemos tirar da história da física, é que toda vez que novos "regimes" experimentais são sondados (por exemplo, altas velocidades, tamanhos pequenos, altas densidades de massa, altas energias), fenômenos são observados que levam a novas teorias (Relatividade Especial, mecânica quântica, Relatividade Geral e o Modelo Padrão, respectivamente)[48].

"Toda vez" é muito forte. Um detalhe, sim, mas também um ponto importante: você não pode simplesmente assumir que qualquer lei em particular falhará em um novo regime. Mas é possível que uma nova lei fundamental esteja envolvida nas estatísticas de Born, e que essa lei se manifeste apenas na vigésima casa decimal em níveis microscópicos (sendo, portanto, indetectável até agora), enquanto se agrega para ter efeitos substanciais em níveis macroscópicos.

Pode haver alguma lei, ainda não descoberta, que faça com que exista apenas um mundo?

Esta é uma noção chocante; implica que todos os nossos gêmeos nos outros mundos - todas as diferentes versões de nós mesmos constantemente se dividindo, não apenas por pesquisadores humanos fazendo medições quânticas, mas por processos entrópicos comuns - na verdade, se foram, deixando-nos sozinhos! Esta versão da Terra seria a única versão que existe no espaço local! Se o cenário inflacionário na cosmologia estiver errado, e a topologia do universo for finita e relativamente pequena - de modo que a Terra não tenha duplicatas distantes que seriam implicadas por um universo exponencialmente vasto - então esta Terra poderia ser a única Terra que existe em qualquer lugar, um pensamento bastante inquietante!

---

48 NT. Texto original em inglês. *If there's one lesson we can take from the history of physics, it's that everytime new experimental "regimes" are probed (e.g. large velocities, small sizes, large mass densities, large energies), phenomena are observed which lead to new theories (Special Relativity, quantum mechanics, General Relativity, and the Standard Model, respectively).*

Mas é perigoso focar demais em hipóteses específicas sobre as quais você não tem razão específica para pensar. Este é o mesmo erro fundamental dos adeptos do Projeto Inteligente, que escolhem qualquer enigma aleatório na genética moderna, e dizem: “Veja, Deus deve ter feito isso!” Por que “Deus”, em vez de um zilhão de outras possíveis explicações? - que você teria pensado muito antes de postular a intervenção divina, se não fosse pelo fato de que você secretamente já começou sabendo a resposta que queria encontrar.

Você não deveria nem perguntar: “Pode haver apenas um mundo?” mas, em vez disso, simplesmente seguir fazendo física, e levantar essa questão particular apenas se novas evidências a exigirem.

Pode haver alguma lei fundamental ainda desconhecida, que dá ao universo um centro privilegiado, que coincidentemente coincide com a Terra - provando assim que Copérnico estava errado o tempo todo, e a Bíblia certa?

Fazer essa pergunta em particular - em vez de um zilhão de outras perguntas nas quais o centro do universo é *Proxima Centauri*, ou o universo tem uma cobertura de pizza favorita e é pepperoni - revela sua agenda oculta. E embora uma pessoa não esclarecida possa não perceber, dar ao universo um centro privilegiado que segue a Terra através do espaço seria bastante difícil de fazer com qualquer lei fundamental matematicamente simples.

O mesmo ocorre ao perguntar se pode haver apenas um mundo. Isso revela um apego sentimental a intuições humanas já provadas erradas. A roda da ciência gira, mas não gira para trás.

Temos razões específicas para ser altamente suspeitos da noção de apenas um mundo. A noção de “um mundo” existe em um nível mais alto de organização, como a localização da Terra no espaço; no nível quântico, não existem limites firmes (embora cérebros que diferem por neurônios inteiros disparando certamente sejam decoerentes). Como uma lei física fundamental identificaria um mundo de alto nível?

Muito pior, qualquer cenário físico no qual houvesse um único mundo sobrevivente, de modo que qualquer medição tivesse apenas um único resultado, violaria a Relatividade Especial.

Conforme a teoria dos muitos mundos, se as mesmas leis se aplicam em todos os níveis, quando você mede um fóton em um par de fótons polarizados emaranhados, acaba em um mundo onde o fóton está polarizado para cima e para baixo, enquanto versões alternativas de você acabam em mundos onde o fóton está polarizado da esquerda para a direita. Da sua perspectiva antes de fazer a medição, as probabilidades são 50/50. A anos-luz de distância, alguém mede o outro fóton em um ângulo de 20° em relação à sua base. Da perspectiva deles, também, a probabilidade de obter qualquer resultado imediato é de 50/50 - eles mantêm um estado invariante de emaranhamento generalizado com sua localização distante, não importa o que você faça. Mas quando vocês dois se encontrarem, anos depois, sua probabilidade de encontrar um amigo que obteve o mesmo resultado é de 11,6%, em vez de 50%.

Se houver apenas um mundo global, então há apenas um único resultado de qualquer medição quântica. Ou você mede o fóton polarizado para cima e para baixo, ou para esquerda e para direita, mas não ambos. A anos-luz de distância, a probabilidade de alguém medir o fóton de maneira semelhante em uma base rotacionada de 20° muda realmente de 50/50 para 11,6%.

Você não pode interpretar isso como um caso de meramente revelarpropriedades que já estavam lá; isso é descartado pelo Teorema de Bell. Não parece haver nenhuma visão consistente do universo em que ambas as medições quânticas tenham um único resultado, e ainda assim ambas as medições sejam predeterminadas, sem que uma influencie a outra. Algo tem que realmente mudar, mais rápido que a luz.

E isso parece uma objeção totalmente geral, não apenas para teorias de colapso, mas para qualquer teoria possível que nos dê um mundo global único! Não há uma visão consistente na qual as medições tenham resultados únicos, mas sejam determinadas localmente (mesmo determinadas localmente de forma aleatória). Alguma influência misteriosa tem que atravessar uma lacuna semelhante a do espaço.

Isso não é uma questão trivial. Você não pode se salvar agitando as mãos e dizendo: “a influência viaja para trás no tempo até a criação dos fótons emaranhados, depois para frente no tempo até o outro fóton, então nunca realmente atravessa uma lacuna semelhante a do espaço.” (Essa visão foi seriamente apresenta-

da, o que lhe dá uma ideia da magnitude do paradoxo implicado por um mundo global único!)

Uma medição tem que mudar a outra, então qual medição acontece primeiro? Existe um espaço global de simultaneidade? Você não pode ter ambas as medições acontecendo "primeiro" porque, sob o Teorema de Bell, não há como a informação local explicar os resultados observados, e assim por diante.

Incidentalmente, esse experimento já foi realizado, e se houver uma influência misteriosa, ela teria que viajar seis milhões de vezes mais rápido que a luz no referencial dos Alpes Suíços. Além disso, a influência misteriosa foi experimentalmente mostrada como não se importando se os dois fótons são medidos em referenciais que fariam com que cada medição ocorresse "antes da outra."

A Relatividade Especial parece contra-intuitiva para nós humanos - como um limite de velocidade arbitrário, que você poderia contornar voltando no tempo e depois avançando novamente. Uma lei da qual você poderia escapar da acusação de violar, se conseguisse esconder seu crime das autoridades.

Mas o que a Relatividade Especial realmente diz é que as intuições humanas sobre espaço e tempo estão simplesmente erradas. Não há um "agora" global, não há "antes" ou "depois" através de intervalos semelhantes ao espaço. A capacidade de visualizar um único mundo global, mesmo em princípio, vem de não entender a Relatividade Especial em um nível visceral. Caso contrário, seria óbvio que a física procede localmente com estados invariantes de emaranhamento distante, e a informação necessária simplesmente não está localmente presente para sustentar um mundo global único.

Pode ser que essa lógica aparentemente impecável seja falha - que minha aplicação do Teorema de Bell e da relatividade para descartar qualquer mundo global único contenha algum pressuposto oculto do qual não estou ciente -

- Mas considere o fardo que uma teoria de mundo único deve agora suportar! Não há absolutamente nenhuma razão em primeiro lugar para suspeitar de um mundo global único; isso simplesmente não é o que a física atual diz! O mundo global único é uma intuição humana antiga refutada, como a ideia de um tempo absoluto universal. O princípio da superposição é visível até mesmo em espelhos semi-refletores; experimentos estão verificando a refutação em níveis crescentes de superposição - mas, acima de tudo, não há mais razão para privilegiar a hipótese da existência de um mundo global único. A escada foi puxada debaixo dessa intuição humana.

Não há evidência experimental de que o mundo macroscópico seja único (já sabemos que o mundo microscópico é superposto). E a perspectiva viola necessariamente a Relatividade Especial, ou dá um salto ainda mais milagroso e viola uma lógica aparentemente impecável. Este último, claro, é muito mais plausível, na prática. Mas isso não é realmente tão plausível em um sentido absoluto. Sem evidência experimental, geralmente é um mau sinal ter que postular milagres lógicos arbitrários.

Quanto ao não-realismo quântico, me parece nada mais do que um Cartão de Saída Livre da Prisão. "Tudo bem violar a Relatividade Especial porque nada disso é real de qualquer maneira!" As equações não podem razoavelmente ser hipotisadas para fornecer previsões tão excelentes sem motivo. O Teorema de Bell descarta a possibilidade óbvia de que a teoria quântica representa um conhecimento imperfeito de algo localmente determinístico.

Além disso, a decoerência macroscópica nos dá uma compreensão perfeitamente realista do que está acontecendo, na qual as equações fornecem boas previsões porque espelham a realidade. Então, a ideia de que as equações quânticas são apenas "sem sentido", e, portanto, é aceitável violar a Relatividade Especial, para podermos ter um mundo global único, não é necessária. Para mim, o não-realismo quântico parece um grande blefe construído em torno de sinais de parada semânticos como "Sem sentido!"

Não é exatamente seguro dizer que a existência de múltiplas Terras é tão bem estabelecida quanto qualquer outra verdade da ciência. A existência de outros mundos quânticos não é tão bem estabelecida quanto a existência de árvores, que a maioria de nós pode observar pessoalmente.

Talvez haja algo na vigésima casa decimal, que se agregue em algo maior em eventos macroscópicos. Talvez haja uma brecha na lógica aparentemente de ferro que diz que qualquer mundo global único deve

violar a Relatividade Especial, porque a informação para sustentar um mundo global único não está localmente disponível. E talvez o Monstro de Espaguete Voador esteja apenas brincando conosco, e o mundo que conhecemos é uma mentira.

Então, tudo o que podemos dizer sobre a existência de múltiplas Terras é que é tão racionalmente provável quanto, por exemplo, a afirmação de que buracos negros em rotação não violam a conservação do momento angular. Temos razões extremamente fundamentais, relacionadas à simetria rotacional do espaço, para suspeitar que a conservação do momento angular está incorporada na natureza subjacente da física. E não temos nenhuma razão específica para suspeitar dessa violação particular de nossas antigas generalizações em um regime de altas energias.

Mas não verificamos a conservação do momento angular para buracos negros giratórios - até onde eu sei. (E como estou falando aqui sobre suposições racionais em estados de conhecimento parcial, o ponto é o mesmo se a observação foi feita e eu ainda não sei). E buracos negros são um regime mais massivo. Portanto, a obediência dos buracos negros não é tão garantida quanto a de que meu vaso sanitário conserva o momento angular ao ser acionado, o que, pensando bem, também não verifiquei…

Ainda assim, se você cometer o erro de pensar demais nessa possibilidade particular, em vez de zilhões de outras possibilidades - e especialmente se você não entender a razão fundamental pela qual o momento angular é conservado - então pode começar a parecer cada vez mais plausível que "buracos negros giratórios violam a conservação do momento angular", à medida que você pensa em mais e mais razões vagamente plausíveis para que isso possa ser verdade.

Mas a probabilidade racional é muito pequena.

Da mesma forma, a probabilidade racional de que exista apenas uma Terra.

Menciono isso para explicar meu hábito de falar como se muitos mundos fosse um fato óbvio. Muitos mundos é um fato óbvio, se você tiver todas as suas ideias alinhadas corretamente (entender física quântica básica, conhecer a teoria formal da probabilidade da Navalha de Ocam, entender a Relatividade Especial, e assim por diante). Na verdade, é consideravelmente mais óbvio para mim do que a proposição de que buracos negros giratórios devem obedecer à conservação do momento angular.

A única razão pela qual a existência de muitos mundos não é universalmente reconhecida como uma previsão óbvia da física que precisa de mágica para ser violada é porque um acidente aleatório na história científica da nossa Terra deu um lugar especial a uma teoria parecida com a teoria do flogisto, que tinha um "colapso" mágico inobservável, mais rápido que a luz, devorando todos os outros mundos. E muitos físicos acadêmicos não tem uma compreensão matemática da Navalha de Ocam, que é o método usual para se livrar de anjos invisíveis na física. Assim, quando se deparam com muitos mundos e isso entra em conflito a intuição deles (prejudicada) de que só existe um mundo, eles dizem: "Ah, isso é multiplicar entidades" — o que é totalmente errado em termos de teoria da probabilidade — e continuam com suas vidinhas normais.

Não estou no meio acadêmico. Não sou obrigado a me curvar a algum físico sênior que não percebeu o óbvio, mas que revisará meus artigos de revista. Não preciso temer ser rejeitado para um cargo vitalício por assustar meus alunos com "contos de ficção científica de outras Terras". Se eu não puder falar claramente, quem poderá?

Então, deixe-me dizer claramente, em nome de todos os físicos que não ousam dizer por si: A teoria dos muitos-mundos vence de forma absoluta, dado nosso estado atual de evidências. Não há mais razão para postular uma única Terra do que há para postular que dois quarks top colidindo se desintegrariam de uma maneira que violaria a Conservação da Energia. Isso requer mais do que uma lei fundamental desconhecida; requer magia.

O debate já deveria ter terminado. Deveria ter terminado há cinquenta anos. O estado da evidência é muito desequilibrado para justificar mais discussões. Não há equilíbrio nesta questão. Não há controvérsia racional a ser ensinada. As leis da teoria da probabilidade são leis, não sugestões; não há flexibilidade na melhor suposição dada essa evidência. Nossos filhos olharão para trás e verão que ainda estávamos discutindo

sobre isso no início do século XXI, e deduzirão corretamente que éramos loucos.

Já envergonhamos nossa Terra o suficiente ao não ver o óbvio. Portanto, pela honra da minha Terra, escrevo como se a existência de muitos mundos fosse um fato estabelecido, porque é. A única questão agora é quanto tempo levará para que as pessoas deste mundo se atualizem.

Parte T - Ciência e Racionalidade

243 - As falhas da ciência antiga

Desta vez, não havia vestes, capuzes ou máscaras. Esperava-se que os estudantes se tornassem amigos e aliados. E todos sabiam por que estavam na sala de aula. Teria sido inútil fingir que não faziam parte da Conspiração.

Seu sensei era Jeffreyssai, que poderia ter sido o melhor de sua era, em sua época. Seus alunos eram os aprendizes mais promissores, ou aqueles em quem os *beisutsukai* viam vantagem política em moldar.

Brennan se enquadrava na última categoria, e sabia disso. Tampouco hesitou em usar o nome de sua Mestra para abrir portas. Você usava todas as vias disponíveis em busca de conhecimento; isso era respeitado aqui.

"-por mais de trinta anos", disse Jeffreyssai. "Nenhum deles viu; nem Einstein, nem Schrödinger, nem mesmo von Neumann." Ele se afastou de seu quadro e virou-se para a turma. "Eu vos pergunto: como eles falharam?"

Os alunos trocaram olhares rápidos, um cálculo de risco mútuo entre os cautelosos e os meramente perplexos. Jeffreyssai era conhecido por seus jogos.

Finalmente, Hiriwa, apelidada de Negra, inclinou-se para frente, tilintando levemente enquanto suas pulseiras entalhadas com equações se moviam em seus tornozelos. "Pelos anos que o sensei mencionou, isso foi duzentos e cinquenta anos após Newton. Certamente, os cientistas daquela época deviam ter compreendido o conceito de uma lei universal."

"Conhecer a lei universal da gravidade", disse o aluno Taji, de um assento próximo, "não é o mesmo que entender o conceito de uma lei universal." Ele era um dos promissores, assim como Hiriwa.

Hiriwa franziu a testa. "Não… foi dito que Newton havia sido elogiado por descobrir a primeira lei universal. Mesmo em sua própria época. Então, era sabido." Hiriwa fez uma pausa. "Mas o próprio Newton já teria partido. Havia alguma injunção religiosa contra propor mais leis universais? Eles se abstiveram por respeito a Newton, ou estavam esperando que seu fantasma falasse? Não estou clara sobre como a ciência antiga era motivada…"

"Não", murmurou Taji, com um riso na voz, "você, realmente, realmente não está."

A expressão de Jeffreyssai era gentil. "Hiriwa, não era religião, nem chumbo na água potável, nem todos tinham Alzheimer, e eles não ficavam sentados o dia todo lendo webcomics. Esqueça o catálogo de horrores dos tempos antigos. Pense apenas em termos de erros cognitivos. O que a ciência antiga poderia estar pensando errado?"

Hiriwa recostou-se com um suspiro. "Sensei, eu realmente não consigo imaginar um erro que faria isso."

"Não seria apenas um erro", Taji a corrigiu. "Como diz o ditado: Erros não viajam sozinhos; eles caçam em matilha."

"Mas a espécie humana inteira?", disse Hiriwa. "Trinta anos?"

“Não era a espécie humana inteira, Hiriwa”, disse Styrlyn. Ele era um dos alunos de aparência mais velha, usando uma barba curta salpicada de grisalho. “Talvez um em cem mil pudesse ter escrito a Equação de Schrödinger de memória. Então, esse teria sido o primeiro e principal erro deles – falha em concentrar suas forças.”

“Poupe-nos da propaganda!” O olhar de Jeffreyssai de repente ficou feroz. “Você não está aqui para proselitismo pela Conspiração Cooperativa, meu senhor político! Não distorça a verdade para defender seus pontos! Acredito que sua Conspiração tenha uma frase: ‘Vantagem comparativa’. Você realmente acha que teria ajudado chamar toda a espécie humana, como existia naquela época, para debater física quântica?”

Styrlyn não se intimidou. “Talvez não, sensei”, ele disse. “Mas se você for comparar aquela época com esta, é uma consideração.”

Jeffreyssai moveu a mão horizontalmente no ar; o gesto de “talvez” que ele usava para descartar um argumento que era verdadeiro, mas irrelevante. “Não é o que eu chamaria de erro principal. O enigma não deveria ter exigido um bilhão de físicos para ser resolvido.”

“Posso pensar em horrores antigos mais específicos”, disse Taji. “Passar o dia todo escrevendo pedidos de subsídios. Ensinar alunos de graduação que prefeririam estar em outro lugar. Precisar publicar trinta artigos por ano para conseguir estabilidade…”

“Mas não estamos falando apenas dos cientistas de status inferior”, disse Yin; ela usava um sorriso levemente provocante. “Dizia-se de Schrödinger que ele se retirou para uma vila por um mês, com sua amante para fornecer inspiração, e emergiu com sua equação epônima. Consideramos um famoso sucesso histórico de nossa metodologia. Alguns físicos antigos entendiam como focar suas energias mentais; e teriam sido experientes o suficiente para fazê-lo, se assim o quisessem.”

“Verdade”, disse Taji. “No final, os encargos administrativos são apenas um obstáculo genérico. Da mesma forma, respostas como: ‘Eles não foram treinados em teoria da probabilidade e não sabiam de vieses cognitivos.’ Nosso sensei parece desejar uma resposta mais específica.”

Jeffreyssai levantou uma sobrancelha encorajadoramente. “Não descarte sua linha de pensamento tão rapidamente, Taji; ela começa a ser relevante. Que tipo de sistema criaria encargos administrativos sobre seu próprio povo?”

“Um sistema que falhou em dar suporte ao seu povo adequadamente”, disse Styrlyn. “Um que falhou em valorizar o trabalho deles.”

“Ah”, disse Jeffreyssai. “Mas há um aluno que ainda não falou. Brennan?”

Brennan não se assustou. Ele deliberadamente esperou tempo suficiente para mostrar que não estava com medo e então disse: “Falta de motivação pragmática, sensei.”

Jeffreyssai sorriu levemente. “Expanda.”

Que tipo de sistema criaria encargos administrativos sobre seu próprio povo?, o sensei lhes havia perguntado. Os outros alunos estavam seguindo suas próprias linhas de pensamento. Brennan, recuando, tinha mais atenção para as poucas dicas de seu professor. Ser o novato nem sempre era uma desvantagem - e ele havia sido ensinado, muito antes de os Bayesianos o acolherem, a aproveitar todas as vantagens disponíveis.

“O Projeto Manhattan”, disse Brennan, “foi lançado com um fim tecnológico específico em mente: uma arma de grande poder, em tempo de guerra. Mas o erro que a Ciência Antiga cometeu em relação à física quântica não teve consequências imediatas para sua tecnologia. Eles estavam confusos, mas não tinham uma necessidade desesperada por uma resposta. Caso contrário, o sistema circundante teria removido todos os obstáculos de seus esforços para resolvê-lo. Certamente o Projeto Manhattan deve ter feito isso - Taji? Você sabe?”

Taji parecia pensativo. “Nem todos os obstáculos - mas tenho certeza de que eles não estavam escre-

vendo pedidos de subsídios no meio do trabalho."

"Então", disse Jeffreyssai. Ele avançou alguns passos, parando diretamente em frente à mesa de Brennan. "Você acha que os cientistas antigos simplesmente não estavam se esforçando o suficiente? Seria porque sua arte não tinha aplicações militares? Um ponto de vista bastante competitivo, eu diria."

"Não necessariamente", disse Brennan com calma. "O pragmatismo também é uma virtude da racionalidade. Um uso desejado para uma teoria quântica teria auxiliado os cientistas antigos de muitas maneiras além de apenas os motivar. Teria dado forma à curiosidade deles e lhes dito o que constituía sucesso ou fracasso."

Jeffreyssai riu levemente. "Não tente adivinhar tanto o que eu gostaria de ouvir, Competidor. Sua primeira declaração chegou mais perto do meu alvo oculto; sua isenção de responsabilidade oh-tão-Bayesiana errou o alvo… O fator que eu tinha em mente, Brennan, era que os cientistas antigos achavam aceitável levar trinta anos para resolver um problema. Todo o seu processo social de ciência era baseado em chegar à verdade eventualmente. Uma teoria errada era descartada eventualmente - assim que a próxima geração de estudantes crescesse familiarizada com a substituta. O trabalho se expande para preencher o tempo alocado, como diz o ditado. Mas as pessoas podem ter pensamentos importantes em muito menos de trinta anos, se esperarem rapidez de si mesmas." Jeffreyssai de repente bateu a mão no braço da cadeira de Brennan. "Quanto tempo você tem para se esquivar de uma faca lançada?"

"Muito pouco tempo, sensei!"

"Menos de um segundo! Dois oponentes estão atacando você! Quanto tempo você tem para adivinhar quem é mais perigoso?"

"Menos de um segundo, sensei!"

"Os dois oponentes se separaram e estão atacando duas de suas namoradas! Quanto tempo você tem para decidir qual você realmente ama?"

"Menos de um segundo, sensei!"

"Um novo argumento mostra que sua preciosa teoria está errada! Quanto tempo você leva para mudar de ideia?"

"Menos de um segundo, sensei!"

"Errado! Não me dê a resposta errada só porque ela se encaixa em um padrão conveniente e eu pareço esperar isso de você! Quanto tempo realmente leva, Brennan?"

O suor estava se formando nas costas de Brennan, mas ele parou e realmente pensou sobre isso…

"Responda, Brennan!"

"Não, sensei! Ainda não terminei de pensar, sensei! Uma resposta seria prematura! Sensei!"

"Ótimo! Continue! Mas não leve trinta anos!"

Brennan respirou fundo, reformulando seus pensamentos. Ele finalmente disse: "Sendo realista, sensei, o melhor cenário seria que eu visse o problema imediatamente; usasse a disciplina de suspender o julgamento; tentasse reavaliar todas as evidências antes de continuar; e, dependendo do quanto eu estivesse emocionalmente apegado à teoria, usasse a técnica de crise de crença para garantir que eu pudesse genuinamente ir para qualquer lado. Então, pelo menos cinco minutos e talvez até uma hora."

"Bom! Você realmente pensou sobre isso desta vez! Pense sobre isso sempre! Quebre padrões! Nos dias da Ciência Antiga, Brennan, não era incomum uma agência de fomento gastar seis meses revisando uma proposta. Eles se permitiam o tempo! Você está sendo avaliado pela sua velocidade, Brennan! A questão não é se você chegará lá eventualmente! Qualquer um pode encontrar a verdade em cinco mil anos! Você precisa

se mover mais rápido!”

“Sim, sensei!”

“Agora, Brennan, você acabou de aprender algo novo?”

“Sim, sensei!”

“Quanto tempo você levou para aprender essa coisa nova?”

Uma escolha arbitrária aí... “Menos de um minuto, sensei, a partir do limite que parece mais óbvio.”

“Menos de um minuto”, repetiu Jeffreyssai. “Então, Brennan, quanto tempo você acha que deveria levar para resolver um grande problema científico, se você não estiver perdendo tempo?”

Agora havia uma pergunta-armadilha, se Brennan já tinha ouvido uma. Não havia como adivinhar que período Jeffreyssai tinha em mente - o que o sensei consideraria muito longo ou muito curto. O que significava que a única saída era apenas tentar a verdade genuína; isso lhe ofereceria a defesa da honestidade, por menor que fosse. “Um ano, sensei?”

“Você acha que poderia ser feito em um mês, Brennan? Em um caso, proponho estipularmos, onde em princípio você já tem evidência experimental suficiente para determinar uma resposta, mas não tanta evidência experimental que você possa se dar ao luxo de cometer erros na interpretação.”

Novamente, nenhuma maneira de adivinhar qual resposta Jeffreyssai poderia querer… “Um mês parece um tempo irrealisticamente curto para mim, sensei.”

“Um tempo curto?”, disse Jeffreyssai, incrédulo. “Quantos minutos em trinta dias? Hiriwa?”

“43.200, sensei”, ela respondeu. “Se você assumir períodos de vigília de dezesseis horas e sono diário, então 28.800 minutos.”

“Assuma, Brennan, que leva cinco minutos inteiros para ter um pensamento original, em vez de aprendê-lo de outra pessoa. Será que mesmo um grande problema científico requer 5.760 insights distintos?”

“Confesso, sensei”, disse Brennan lentamente, “que nunca pensei nisso dessa maneira antes... mas você me diz que esse é realmente um nível realista de produtividade?”

“Não”, disse Jeffreyssai, “mas também não é realista pensar que um único problema requer 5.760 insights. E sim, isso já foi feito.”

Jeffreyssai deu um passo para trás e sorriu benevolentemente. Todos os alunos na sala se enrijeceram; eles conheciam aquele sorriso. “Embora nenhum de vocês tenha acertado a resposta particular que eu tinha em mente, suas respostas foram tão razoáveis quanto a minha. Exceto a de Styrlyn, receio. Até a resposta de Hiriwa não estava totalmente errada: a tarefa de propor novas teorias já foi considerada um dever sagrado reservado para aqueles de alto status, havendo um suprimento limitado de problemas em circulação, naquela época. Mas a resposta de Brennan é particularmente interessante, e estou inclinado a testar sua teoria da motivação.”

Oh, inferno, Brennan disse silenciosamente para ele mesmo. Jeffreyssai estava gesticulando para Brennan se levantar diante da classe.

Quando Brennan se levantou, Jeffreyssai se sentou na cadeira de Brennan.

“Brennan-sensei”, disse Jeffreyssai, “você tem cinco minutos para pensar em algo incrivelmente brilhante para dizer sobre o fracasso da ciência antiga na física quântica. Quanto ao resto de nós, nosso trabalho será olhar para você com expectativa. Só posso imaginar como será constrangedor se você não conseguir pensar em nada de bom.”

Bastardo. Brennan não disse em voz alta. O rosto de Taji mostrava um pouco de simpatia; Styrlyn se mantinha distante do jogo; mas Yin estava olhando para ele com interesse sardônico. Pior, Hiriwa estava olhando para ele com expectativa, assumindo que ele aceitaria o desafio. E Jeffreyssai estava boquiaberto, esperando as palavras de sabedoria do guru. Dane-se, sensei.

Brennan não entrou em pânico. Estava muito, muito, muito longe de ser a situação mais assustadora que ele já havia enfrentado. Ele tomou um momento para decidir como pensar; então pensou.

Aos quatro minutos e trinta segundos, Brennan falou. (Havia uma arte em tais coisas; já que você estava fazendo isso de qualquer maneira, você poderia muito bem fazer parecer fácil.)

"Uma mulher sábia", disse Brennan, "me disse uma vez que é mais sábio considerar nossos "eus" do passado como tolos além da redenção - ver as pessoas que já fomos como idiotas completos. Eu não necessariamente digo isso; mas foi o que ela me disse, e há mais do que um grão de verdade nisso. Enquanto estivermos inventando desculpas para o passado, tentando fazê-lo parecer melhor, respeitando-o, não podemos romper com ele. Ocorre-me que a regra pode não ser diferente para civilizações humanas. Então, tentei olhar para trás e considerar os cientistas antigos como simples tolos."

"O que eles não eram", disse Jeffreyssai.

"O que eles não eram", continuou Brennan. "Em termos de inteligência bruta, eles sem dúvida me superavam. Mas me ocorreu que a dificuldade em ver o que os cientistas antigos fizeram de errado pode ter sido em respeitar demais os nomes antigos e lendários. E isso realmente produziu um insight."

"Chega de introdução, Brennan", disse Jeffreyssai. "Se você encontrou um insight, declare-o."

"Os cientistas antigos não foram treinados..." Brennan fez uma pausa. "Não, 'não treinados' não é o conceito. Eles foram treinados para a tarefa errada. Naquela época, não havia Conspirações, nem verdades secretas; assim que os cientistas antigos resolviam um grande problema, eles publicavam a solução para o mundo e uns para os outros. Problemas em aberto verdadeiramente assustadores e confusos teriam sido extremamente raros e usados no momento em que fossem resolvidos. Portanto, não teria sido possível treinar pesquisadores antigos para trazer ordem ao caos científico. Eles teriam sido treinados para outra coisa - não tenho certeza do quê..."

"Treinados para manipular qualquer ciência que já tivesse sido descoberta", disse Taji. "Era uma tarefa difícil o suficiente para os professores da Ciência Antiga treinarem seus alunos a usar o conhecimento existente ou seguir metodologias já conhecidas; isso era tudo o que os professores da Ciência Antiga aspiravam transmitir."

Brennan assentiu. "O que é muito diferente de criar uma nova ciência própria. Os cientistas antigos, confrontados com problemas da teoria quântica, podem nunca ter enfrentado esse tipo de medo antes - o desânimo de não saber. Os cientistas antigos podem ter se apegado a respostas insatisfatórias prematuramente, porque estavam acostumados a trabalhar com um corpo de conhecimento organizado e consensual."

"Bom, Brennan", murmurou Jeffreyssai.

"Mas acima de tudo", continuou Brennan, "um cientista antigo não poderia ter praticado o problema real que os cientistas quânticos enfrentaram - o de resolver uma grande confusão. Era algo que você fazia uma vez na vida, se tivesse sorte, e, como Hiriwa observou, Newton não estaria mais por perto. Então, embora os físicos antigos que erraram a teoria quântica não fossem pouco inteligentes, eles eram, em um sentido forte, amadores - improvisando todo o processo de mudança de paradigma."

"E nenhuma teoria da probabilidade", observou Hiriwa. "Então, qualquer um que tivesse sucesso no problema não teria ideia do que havia acabado de fazer. Eles seriam incapazes de comunicar isso a mais ninguém, exceto vagamente."

"Sim", disse Styrlyn. "E era apenas um punhado de pessoas que podiam abordar o problema, sem treinamento para fazê-lo; esses são os físicos cujos nomes foram transmitidos para nós. Um punhado de pes-

soas, fazendo um punhado de descobertas cada. Não teria sido suficiente para sustentar uma comunidade. Cada cientista antigo abordando uma nova mudança de paradigma teria precisado redescobrir as regras do zero."

Jeffreyssai levantou-se da mesa de Brennan. "Aceitável, Brennan; você me surpreende, na verdade. Terei que pensar mais sobre esse seu método." Jeffreyssai foi até a porta da sala de aula, depois olhou para trás. "No entanto, eu tinha em mente pelo menos outra falha importante da ciência antiga, que nenhum de vocês sugeriu. Espero receber uma lista de possíveis falhas amanhã. Espero que a falha que tenho em mente esteja na lista. Vocês têm 480 minutos, excluindo o tempo de sono. Vejo cinco de vocês aqui. O desafio não requer mais de 480 insights para resolver, nem mais de 96 insights em série."

E Jeffreyssai saiu da sala.

244 - O Dilema: Ciência ou Bayes?

Eli: Você tem escrito muito sobre física recentemente. Por quê?

— Shane Legg (e várias outras pessoas)

> Considerando sua explicação sobre MQ, que para mim soa perfeitamente lógica, parece óbvio e normal que muitos mundos seja esmagadoramente provável. Só parece bom demais para ser verdade que eu agora entenda o que muitos físicos geniais de quântica ainda não entendem. […] Claro que posso explicar tudo isso, e ainda acho que você está certo, só estou desconfiado de mim mesmo por acreditar na primeira explicação plausível que encontrei.
>
> — *Recovering_irrationalist*

Recovering_irrationalist, você não tem ideia de como fiquei feliz em ver seu comentário.

É claro que eu tinha mais de um motivo para passar todo aquele tempo escrevendo sobre física quântica. Gosto de ter muitos motivos ocultos. É o mais próximo que posso chegar eticamente de ser um supervilão.

Mas para dar um exemplo de um propósito que eu só poderia alcançar discutindo física quântica...

Na física, você pode ter questões absolutamente claras. Não no sentido de que as questões sejam triviais de explicar. Mas se você tentar aplicar Bayes à saúde ou economia, pode não conseguir estabelecer formalmente qual é a hipótese mais simples, ou o que as evidências apoiam. Mas quando digo que "a decoerência macroscópica é mais simples que o colapso", é realmente uma simplicidade estrita; você poderia escrever as duas hipóteses como programas e contar as linhas de código. Nem a própria evidência está em disputa.

Eu queria um exemplo muito claro - Bayes diz "zigue", isso é um zague - quando chegasse a hora de romper sua lealdade à Ciência.

"Ah, claro," você diz, "os físicos estragaram a coisa dos muitos mundos, mas dê um desconto a eles, Eliezer! Ninguém jamais afirmou que o processo social da ciência era perfeito. As pessoas são humanas; elas cometem erros."

Mas os físicos que se recusam a adotar muitos mundos não estão desobedecendo às regras da Ciência. Eles estão obedecendo às regras da Ciência.

A tradição passada através das gerações diz que uma nova teoria física surge com novas previsões experimentais que a distinguem da teoria antiga. Você realiza o teste, e a nova teoria é confirmada ou falsificada. Se for confirmada, você faz uma grande celebração, chama os jornais e distribui Prêmios Nobel para todos; quaisquer velhos professores eméritos que se recusem a se converter são discretamente tolerados. Se a teoria for refutada, o principal proponente se retrata publicamente e ganha uma reputação de honestidade.

Não é assim que as coisas funcionam na ciência; é assim que as coisas deveriam funcionar na Ciência. É o ideal ao qual todos os bons cientistas aspiram.

Agora, muitos mundos surge, e não parece fazer nenhuma nova previsão em relação à teoria antiga. Isso é suspeito. E há todos esses outros mundos, mas você não pode vê-los. Isso é realmente suspeito. Simplesmente não parece científico.

Se você chegou tão longe quanto Recovering_irrationalist — de modo que muitos mundos agora parecem perfeitamente lógicos, óbvios e normais — e você também começou como um Racionalista Tradicional, então você deveria conseguir alternar entre a visão Científica e a visão Bayesiana, como um Cubo de Necker.

Então agora coloque seus Óculos de Ciência — você ainda os tem por aí em algum lugar, certo? Esqueça tudo o que sabe sobre complexidade de Kolmogorov, indução de Solomonoff ou Comprimentos Mínimos de Mensagem. Isso não faz parte do treinamento tradicional. Você apenas observa algo para ver quão "simples" parece. A palavra "testável" não evoca uma imagem mental do Teorema de Bayes governando fluxos de probabilidade; evoca uma imagem mental de estar em um laboratório, realizando um experimento e tendo a celebração (ou retratação pública) depois.

Com os Óculos de Ciência: A teoria quântica atual passou em todos os testes experimentais até agora. Muitos mundos não fazem nenhuma nova previsão testável — os novos fenômenos incríveis que prevê estão todos escondidos onde não podemos vê-los. Você pode se virar bem sem supor os outros mundos, e é exatamente isso que deve fazer. Tudo isso cheira a ficção científica. Mas deve-se admitir que a física quântica é uma questão muito profunda e muito confusa, e quem sabe quais descobertas podem estar por vir? Me avise quando a teoria de muitos mundos fizer uma previsão testável.

Tire os Óculos de Ciência, coloque os Óculos de Bayes de volta.

Com os Óculos de Bayes: As equações quânticas mais simples que cobrem todas as evidências conhecidas não têm uma exceção especial para massas de tamanho humano. Nem há sequer motivo para fazer essa pergunta em particular. Próximo!

Okay, então isso é um problema que podemos consertar em cinco minutos com fita adesiva e cola?

Não.

Hã? Por que não simplesmente ensinar às novas turmas de cientistas que se formam sobre indução de Solomonoff e a Regra de Bayes?

Séculos atrás, havia uma ideia generalizada de que os Sábios podiam desvendar os segredos do universo apenas pensando sobre eles, enquanto sair e olhar para as coisas era menor, inferior, ingênuo e acabaria por enganá-lo. Você não podia confiar na aparência das coisas — apenas o pensamento poderia ser seu guia.

A ciência começou como uma rebelião contra essa Sabedoria Profunda. No cerne está a crença pragmática de que os seres humanos, sentados em suas poltronas tentando ser Profundamente Sábios, simplesmente derivam para o país do nunca. Você não podia confiar em seus pensamentos. Você tinha que fazer previsões experimentais antecipadas — previsões que ninguém mais havia feito antes — realizar o teste e confirmar o resultado. Isso era evidência. Sentar-se em sua poltrona, pensando sobre o que parecia razoável… não seria considerado para prejudicar sua teoria, porque a Ciência não era uma crença idealista sobre pragmatismo, ou sujar as mãos. Era, ao contrário, o ditame de que apenas o experimento decidiria. Apenas experimentos poderiam julgar sua teoria — não sua nacionalidade, ou suas profissões religiosas, ou o fato de que você havia inventado a teoria em sua poltrona. Apenas experimentos! Se você se sentasse em sua poltrona e chegasse a uma teoria que fizesse uma previsão nova, e o experimento confirmasse a previsão, então nos importaríamos com o resultado do experimento, não com a origem de sua hipótese.

Isso é Ciência. E se você disser que a teoria dos muitos mundos deve substituir a Interpretação de Copenhague, que é imensamente bem-sucedida, acrescentando todas essas Terras gêmeas que não podem ser observadas, só porque ela parece mais legal e elegante - não porque ela esmagou a teoria antiga com uma previsão experimental superior - então você está ignorando a regra científica central que impede que as pessoas saiam correndo e coloquem anjos em todas as teorias, porque os anjos parecem mais razoáveis e elegantes.

Você acha que ensinar algumas pessoas sobre indução de Solomonoff resolverá esse problema? O laureado com o Nobel Robert Aumann - que primeiro provou que agentes bayesianos com prioris semelhantes não podem concordar em discordar - é um judeu ortodoxo praticante. Aumann ajudou um projeto para testar a Torá em busca de "códigos bíblicos", profecias ocultas de Deus - e concluiu que o projeto havia falhado em confirmar a existência dos códigos. Você quer que Aumann pense que uma vez que você tem indução de Solomonoff, pode esquecer o método experimental? Você acha que isso vai ajudá-lo? E a maioria dos cientistas por aí não chegará ao nível de Robert Aumann.

Okay, agora coloque os Óculos de Bayes de volta. Você realmente vai acreditar que grandes partes da função de onda desaparecem quando você não pode mais as ver? Como resultado do único fenômeno não-linear não-unitário não-diferenciável não-simétrico-CPT acausal mais rápido que a luz informalmente especificado em toda a física? Apenas porque, por pura contingência histórica, a versão estúpida da teoria foi proposta primeiro?

Você fará uma grande modificação em um modelo científico e acreditar em zilhões de outros mundos que você não pode ver, sem um momento definidor de triunfo experimental sobre o modelo antigo?

Ou você rejeitará a teoria da probabilidade?

Você dará sua lealdade à Ciência ou a Bayes?

Michael Vassar uma vez observou (em tom de brincadeira) que era bom que a maioria da espécie humana acreditasse em Deus, porque caso contrário, ele teria muita dificuldade em rejeitar o majoritarismo. Mas como a opinião majoritária de que Deus existe é simplesmente inacreditável, não temos escolha senão rejeitar os argumentos filosóficos extremamente fortes a favor do majoritarismo.

Você pode ver (uma das razões) por que fui a tais extremos para explicar a teoria quântica. Aqueles bons em matemática agora devem conseguir visualizar tanto a decoerência macroscópica quanto a teoria da probabilidade da simplicidade e testabilidade - entender a insanidade de um único mundo global em um nível visceral.

Eu queria apresentar a você um dilema agradável e nítido entre rejeitar o método científico ou abraçar a insanidade.

Por quê? Vou te dar uma dica: Não é só porque sou mau. Se você quiser adivinhar meus motivos aqui, pense além da primeira resposta óbvia.

PS: Se você tentar encontrar maneiras inteligentes de escapar do dilema, você só será derrubado em ensaios futuros. Você foi avisado.

245 - A ciência não confia na sua racionalidade

Scott Aaronson sugere que muitos mundos e libertarianismo são semelhantes. Ambos são casos de engolir a bala, em vez de esquivá-la:

> Libertarianismo e MWI (Teoria dos Muitos Mundos) são ambas grandes teorias filosóficas que partem de premissas que quase todas as pessoas instruídas aceitam (mecânica quântica em um caso, Economia 101 no outro). Elas afirmam chegar a conclusões que a maioria das pessoas instruídas rejeita, ou pelo menos fica perplexa (a existência de universos paralelos/a conveniência de eliminar os corpos de bombeiros)[49].

Essa é uma analogia que nunca teria me ocorrido.

Já argumentei antes que a Ciência rejeita a MWI, mas Bayes a aceita. (Aqui, "Ciência" está em maiúscula porque falamos da forma idealizada da Ciência, não apenas do processo social real da ciência.)

Além disso, me parece que há uma analogia profunda entre o libertarianismo (com "l" minúsculo) e a Ciência:

1. Ambos se baseiam em uma desconfiança pragmática de argumentos aparentemente razoáveis.
2. Ambos tentam construir sistemas mais confiáveis do que as pessoas neles.
3. Ambos aceitam que as pessoas são falhas e tentam usar essas falhas para impulsionar o sistema.

O argumento central do libertarianismo é a desconfiança historicamente motivada de teorias adoráveis sobre "Como a sociedade seria muito melhor se apenas fizéssemos uma regra dizendo XYZ". Se esse tipo de truque realmente funcionasse, mais regulamentações se correlacionariam com maior crescimento econômico à medida que a sociedade passasse de ótimos locais para globais. Mas quando alguma pessoa ou grupo de interesse ganha poder suficiente para começar a fazer tudo o que acha ser uma boa ideia, a história nos mostra que o que realmente acontece é a França Revolucionária ou a Rússia Soviética.

Os planos que, em teoria, deveriam ter deixado todos felizes para sempre não têm os resultados previstos por argumentos aparentemente razoáveis. E o poder corrompe e atrai os corruptos.

Então, você regula o mínimo possível, porque não pode confiar nas teorias adoráveis nem nas pessoas que as implementam.

Você não repreende as pessoas por serem egoístas. Você tenta construir um sistema eficiente de produção com participantes egoístas, exigindo que as transações sejam voluntárias. Assim, as pessoas são forçadas a jogar jogos de soma positiva, porque é assim que elas conseguem que a outra parte assine o contrato. Com a violência contida e os contratos aplicados, o egoísmo individual pode impulsionar um sistema globalmente produtivo.

Claro que nada disso funciona tão bem, na prática, quanto na teoria, e não entrarei em falhas de mercado, problemas de bens comuns, etc. O argumento central do libertarianismo não é que ele funcionaria

49 NT. Texto original em inglês. *Libertarianism and MWI are both grand philosophical theories that start from premises that almost all educated people accept (quantum mechanics in the one case, Econ 101 in the other), and claim to reach conclusions that most educated people reject, or are at least puzzled by (the existence of parallel universes / the desirability of eliminating fire departments).*

em um mundo perfeito, mas que se degrada graciosamente na realidade. Ou melhor, se degrada menos desajeitadamente do que qualquer outro princípio econômico conhecido. (Pessoas que veem o Libertarianismo como a solução perfeita para pessoas perfeitas me parecem meio que perdendo o ponto da coisa da "desconfiança pragmática".)

A Ciência primeiro se conheceu como uma rebelião contra confiar na palavra de Aristóteles. Se as pessoas daquela revolução tivessem simplesmente dito: "Vamos confiar em nós mesmos, não em Aristóteles!", elas teriam brilhado e desaparecido como a Revolução Francesa.

Mas a Revolução Científica durou porque — como a Revolução Americana — os arquitetos propuseram uma filosofia mais estranha: "Não confiemos em ninguém! Nem mesmo em nós mesmos!"

No início, surgiu a ideia de que não podemos simplesmente descartar o raciocínio de poltrona de Aristóteles e substituí-lo por um raciocínio de poltrona diferente. Precisamos falar com a Natureza e ouvir realmente o que Ela diz em resposta. Isso, por si só, foi um golpe de gênio.

Então veio o desafio da implementação. As pessoas são teimosas e podem não querer aceitar o veredicto do experimento. Devemos balançar um dedo desaprovador para elas e dizer "Malcriados"?

Não; assumimos e aceitamos que cada cientista individual pode estar loucamente apegado às suas teorias pessoais. Nem assumimos que alguém possa ser treinado para sair dessa tendência — não tentamos escolher Juízes Eminentes que supostamente sejam imparciais.

Em vez disso, tentamos aproveitar o desejo teimoso do cientista individual de provar sua teoria pessoal, dizendo: "Faça uma nova previsão experimental e faça o experimento. Se você estiver certo e o experimento for replicado, você ganha."Enquanto os cientistas acreditarem que isso é verdade, eles têm um motivo para fazer experimentos que podem falsificar suas próprias teorias. Apenas aceitando a possibilidade de derrota é possível vencer. E qualquer grande afirmação exigirá replicação; isso dá aos cientistas um motivo para serem honestos, sob pena de grande constrangimento.

E assim, a teimosia dos cientistas individuais é aproveitada para produzir um fluxo constante de conhecimento no nível do grupo. O Sistema é um pouco mais confiável que suas partes.

O libertarianismo depende secretamente de que a maioria dos indivíduos seja pró-social o suficiente para dar gorjeta em um restaurante que nunca mais visitarão. Uma economia de agentes genuinamente egoístas de nível humano implodiria. Da mesma forma, a Ciência depende de que a maioria dos cientistas não cometa pecados tão flagrantes que não possam racionalizá-los.

Na medida em que os cientistas acreditam que podem promover suas teorias jogando política acadêmica - ou manipular os métodos estatísticos para potencialmente ganhar sem chance de perder - ou na medida em que ninguém se preocupa em replicar as afirmações - a ciência se degrada em eficácia. Mas ela se degrada graciosamente, considerando como essas coisas funcionam.

A parte na qual as previsões bem-sucedidas pertencem à teoria e aos teóricos que originalmente as fizeram, e não podem simplesmente ser roubadas por uma teoria que surge depois — sem uma nova previsão experimental — é uma característica importante desse processo social.

O resultado é que a Ciência não é facilmente reconciliada com a teoria da probabilidade. Se você fizer um cálculo de teoria da probabilidade corretamente, você obterá a resposta racional. A Ciência não confia na sua racionalidade e não depende da sua capacidade de usar a teoria da probabilidade como árbitro da verdade. Ela quer que você estabeleça um experimento definitivo.

Considerar a Ciência como uma mera aproximação de algum ideal de racionalidade baseado em teoria da probabilidade… certamente pareceria racional. Parece haver um argumento extremamente razoável de que o Teorema de Bayes é a estrutura oculta que explica por que a Ciência funciona. Mas subordinar a Ciência ao grande esquema do Bayesianismo, e deixar o Bayesianismo entrar e anular o veredicto da Ciência quando isso parecer apropriado, não é um passo trivial!

A Ciência é construída em torno da suposição de que você é estúpido e autoilusório demais para simplesmente usar a indução de Solomonoff. Afinal, se fosse tão simples, não precisaríamos de um processo social de ciência… certo?

Então, você vai acreditar em fadas quânticas de "colapso" mais rápidas que a luz, afinal? Ou você acha que é mais esperto que isso?

246 - Quando a ciência não pode ajudar

Era uma vez, um Eliezer mais jovem tinha uma teoria boba. Digamos que a teoria boba do Eliezer aos 18 anos era que a consciência era causada por curvas fechadas semelhantes ao tempo escondidas na gravidade quântica. Isso não é a história completa, nem perto disso, mas serve como ponto de partida.

E chegou um momento em que olhei para trás e percebi:

1. Eu havia seguido cuidadosamente tudo o que me disseram ser tradicionalmente racional, no curso de me perder. Por exemplo, eu tinha cuidado para acreditar apenas em teorias bobas que fizessem novas previsões experimentais, por exemplo, que os microtúbulos neurais seriam encontrados para suportar estados quânticos coerentes.
2. A ciência teria ficado perfeitamente bem com meu gasto de dez anos tentando testar minha teoria boba, apenas para obter um resultado experimental negativo, contanto que eu então dissesse: “Oh, bem, acho que minha teoria estava errada.”

Da perspectiva da Ciência, é assim que as coisas deveriam funcionar - diversão para todos. Você admitiu seu erro! Bom pra você! Não é disso que se trata a Ciência?

Mas e se eu não quisesse desperdiçar dez anos?

Bem… A Ciência não tinha muito a dizer sobre isso. Como a Ciência poderia dizer qual teoria estava certa, antecipadamente ao teste experimental? A Ciência não se importa de onde sua teoria vem - ela apenas diz: “Vá testá-la.”

Esta é a grande força da Ciência, e também sua grande fraqueza.

Gray Area perguntou:

Eliezer, por que você está preocupado com questões intestáveis?

Porque questões que podem ser facilmente testadas imediatamente são difíceis para a Ciência errar.

Quero dizer, claro, quando já há evidência experimental definitiva e inequívoca disponível, vá em frente. Por que diabos você não faria isso?

Mas às vezes uma questão terá consequências experimentais muito grandes e muito definidas em seu futuro - porém você não pode testá-la experimentalmente facilmente agora - e ainda assim há um forte argumento racional.

Superposições quânticas macroscópicas são facilmente testáveis: bastaria precisão nanotecnológica, temperaturas muito baixas e uma área clara do espaço interestelar. Ah, claro, você não pode fazer isso agora, porque é muito caro ou impossível para a tecnologia de hoje ou algo assim - mas em teoria, claro! Por que, talvez um dia eles vão executar civilizações inteiras em computadores quânticos macroscopicamente superpostos, bem longe em um volume bem varrido de um Grande Vazio. (Perguntar o que o não-realismo quântico diz sobre o status de quaisquer observadores dentro desses computadores ajuda a revelar a subespecificação do não-realismo quântico.)

Isso não parece imediatamente relevante para sua vida, eu acho, mas estabelece o padrão: nem tudo com consequências futuras é barato para testar agora.

A psicologia evolutiva é outro exemplo de um caso em que a racionalidade tem que assumir o controle da ciência. Embora as teorias da psicologia evolutiva formem um todo conectado, apenas algumas dessas teorias são facilmente testáveis experimentalmente. Mas você ainda precisa das outras partes da teoria, porque elas formam uma teia conectada que ajuda você a formular as hipóteses que são realmente testáveis - e então as hipóteses auxiliares são sustentadas em um sentido bayesiano, mas não sustentadas experimentalmente. A ciência proferiria um veredito de "não comprovado" em partes individuais de uma malha teórica conectada que é experimentalmente produtiva na totalidade. Precisaríamos de um novo tipo de veredito para isso, algo como "indiretamente sustentado".

Ou que tal a criônica?

A criônica é um exemplo arquetípico de uma questão extremamente importante (150.000 pessoas morrem por dia) que terá consequências enormes no futuro previsível, mas não oferece evidência experimental definitiva e inequívoca que possamos obter agora.

Então você diz: "Eu não acredito em criônica porque não foi comprovada experimentalmente, e você não deve acreditar em coisas que não foram comprovadas experimentalmente"?

Bem, de uma perspectiva bayesiana, isso está incorreto. Ausência de evidência é evidência de ausência apenas no grau em que poderíamos razoavelmente esperar que a evidência aparecesse. Se alguém está alardeando que o óleo de cobra cura o câncer, é razoável esperar que, se o óleo de cobra estivesse realmente curando o câncer, algum cientista estaria realizando um estudo controlado para verificar isso - que, no mínimo, os médicos estariam relatando estudos de casos de recuperações surpreendentes - e, portanto, a ausência dessa evidência é uma forte evidência de ausência. Mas as "lacunas no registro fóssil" não são uma forte evidência contra a evolução; os fósseis são formados apenas raramente e, mesmo que uma espécie intermediária tenha de fato existido, não seria muito provável que ela fosse fossilizada e encontrada.

Reviver um mamífero criogênicamente congelado simplesmente não é nada que você esperaria poder fazer com a tecnologia moderna, mesmo que as nanotecnologias futuras possam de fato realizar uma ressuscitação bem-sucedida. É assim que vejo Bayes vendo isso.

Ah, e quanto aos argumentos reais para a criônica - eu não vou entrar neles no momento. Mas se você seguiu a física e as sequências anti-zumbi, deve parecer muito mais plausível agora que qualquer coisa que preserve o padrão das sinapses preserve o máximo de "você" que é preservado de uma noite de sono até o despertar da manhã.

Agora, para ser justo, alguém que diz: "Eu não acredito em criônica porque não foi comprovada experimentalmente" está aplicando erroneamente as regras da Ciência; este não é um caso em que a ciência dá realmente a resposta errada. Na ausência de um teste experimental definitivo, o veredito da ciência aqui é "Não comprovado". Qualquer pessoa que interprete isso como uma rejeição está dando um passo extra fora da ciência, não um passo em falso na ciência.

A página Wikiquotes de John McCarthy o tem dizendo: "Suas afirmações equivalem a dizer que se a IA for possível, ela deve ser fácil. Por que isso?" [1] A página Wikiquotes não diz a que McCarthy estava respondendo, mas eu poderia arriscar um palpite.

O erro geral provavelmente surge porque há casos em que a ausência de prova científica é forte evidência - porque um experimento seria facilmente realizável, então a falha em realizá-lo é em si suspeita. (Embora não tão suspeito quanto eu costumava pensar - com todas as estranhas evidências anedóticas variadas chegando de fontes respeitáveis, por que diabos ninguém está testando a teoria de Seth Roberts sobre a supressão do apetite? [2])

Outro fator de confusão pode ser que se você testar o Medicamento X em 1.000 sujeitos e descobrir que 56% do grupo de controle e 57% do grupo experimental se recuperam, algumas pessoas chamarão isso de um veredito de "Não comprovado". Eu chamaria de um veredito experimental de "Medicamento X não

funciona bem, se é que funciona". Apenas porque esse veredito é teoricamente retratável diante de novas evidências, não o torna ambíguo.

Em todo caso, no momento você tem pessoas descartando a criônica imediatamente como "não científica", como se fosse algum tipo de medicamento que você poderia administrar facilmente a 1.000 pacientes e ver o que aconteceria. "Me chame quando os crionicistas realmente reviverem alguém", dizem eles; o que, como Mike Li observa, é como dizer: "Eu me recuso a entrar nesta ambulância; me chame quando ela já estiver no hospital." Talvez Martin Gardner os tenha alertado contra acreditar em coisas estranhas sem evidências experimentais. Então eles esperam pelo veredito definitivo e inequívoco da Ciência, enquanto sua família, amigos e 150.000 pessoas por dia estão morrendo agora, e podem ou não ser salvos -

- uma aposta calculada que você só poderia fazer racionalmente.

A força motriz da Ciência é obter uma montanha de evidências tão grande que nem mesmo cientistas humanos falíveis possam interpretá-la erroneamente. Mas mesmo isso às vezes dá errado, quando as pessoas ficam confusas sobre qual teoria prevê o quê, ou assam componentes extremamente difíceis de testar em uma versão inicial de sua teoria. E às vezes você simplesmente não consegue obter nenhuma evidência experimental clara.

De qualquer forma, você tem que tentar fazer a coisa que a Ciência não confia em ninguém para fazer - pensar racionalmente e descobrir a resposta antes de ser atingido por ela.

(Ah, e às vezes um resultado experimental "desconfirmatório" parece: "Sua espécie inteira acabou de ser extinta! Você agora está cientificamente obrigado a renunciar à sua teoria. Se você se retratar publicamente, muito bem! Lembre-se, é preciso uma mente forte para abrir mão de crenças fortemente arraigadas. Sinta-se à vontade para tentar outra hipótese da próxima vez!")

Referências

[1] Não mais no Wikiquotes, mas incluído na página de citações pessoais de McCarthy.

[2] Seth Roberts, "What Makes Food Fattening?: A Pavlovian Theory of Weight Control" (Manuscrito não publicado, 2005), http://media.sethroberts.net/about/whatmakesfoodfattening.pdf.

247 - A ciência não é rigorosa o suficiente

Era uma vez um Eliezer mais jovem com uma teoria estúpida. O Eliezer aos 18 anos, teve o cuidado de seguir os preceitos da Racionalidade Tradicional que lhe haviam sido ensinados; ele se certificou de que sua teoria estúpida tivesse consequências experimentais. O Eliezer aos 18 anos professou, segundo as virtudes de um cientista que lhe haviam sido ensinadas, que desejava testar sua teoria estúpida.

Isso era tudo o que era necessário para ser virtuoso, conforme o que o Eliezer aos 18 anos havia aprendido ser virtude no caminho da ciência.

Não era nem remotamente a ordem de esforço que teria sido necessário para acertar.

Os ideais tradicionais da Ciência concedem estrelas douradas com muita facilidade. Resultados experimentais negativos também são conhecimentos, então todos que jogam ganham um prêmio. Contanto que você possa pensar em algum tipo de experimento que teste sua teoria, e você faça o experimento, e aceite os resultados, você jogou pelas regras; você é um bom cientista.

Você não necessariamente acertou, mas é um bom cidadão cumpridor da ciência.

(Observo neste ponto que estou falando de Ciência, não do processo social da ciência como ela realmente funciona na prática, por dois motivos. Primeiro, me desviei ao tentar seguir o ideal da Ciência - não fui rejeitado por um editor de revista com rancor, e não estava tentando imitar as falhas do meio acadêmico. Segundo, se eu apontar um problema com o ideal como ele é tradicionalmente pregado, os cientistas reais não são forçados a se desviarem da mesma forma!)

A ciência começou como uma rebelião contra grandes esquemas filosóficos e raciocínio de poltrona. Portanto, a Ciência não inclui uma regra sobre quais tipos de hipóteses você pode ou não testar; isso é deixado para o cientista individual. Tentar adivinhar isso a priori exigiria algum tipo de grande esquema filosófico e raciocínio antes da evidência. Como um ideal social, a Ciência não o julga como uma pessoa má por apresentar hipóteses heréticas; experimentos honestos e a aceitação dos resultados são virtude para um cientista.

Enquanto a maioria dos cientistas conseguir aceitar evidências experimentais definitivas, inconfundíveis e inequívocas, a ciência pode progredir. Pode acontecer muito lentamente - pode levar mais tempo do que deveria - você pode ter que esperar que uma geração de anciãos morra - mas, eventualmente, a catraca do conhecimento avança mais um degrau. Ano após ano, década após década, a roda gira para frente. É o suficiente para sustentar uma civilização.

Então, isso é tudo o que a Ciência pede realmente de você - a capacidade de aceitar a realidade quando você é golpeado na cabeça com ela. Não é muito, mas é o suficiente para sustentar uma cultura científica.

Contraste isso com a noção que temos na teoria da probabilidade, de um julgamento racional quantitativo exato. Se 1% das mulheres que se apresentam para um exame de rotina têm câncer de mama, e 80% das mulheres com câncer de mama recebem mamografias positivas, e 10% das mulheres sem câncer de mama recebem falsos positivos, qual é a probabilidade de uma mulher com exame de rotina e mamografia positiva ter câncer de mama? É 7,5%. Você não pode dizer: “Acredito que ela não tem câncer de mama, porque o experimento não é definitivo o suficiente.” Você não pode dizer: “Acredito que ela tem câncer de mama, porque é sábio ser pessimista e isso é o que o único experimento até agora parece indicar.” Sete vír-

gula cinco por cento é a estimativa racional dada esta evidência, não 7,4% ou 7,6%. As leis da probabilidade são leis.

Está escrito nas Doze Virtudes, na terceira virtude, leveza:

> Se você considera a evidência como uma restrição e procura se libertar, você se vende às correntes de seus caprichos. Pois você não pode fazer um mapa verdadeiro de uma cidade sentando em seu quarto com os olhos fechados e desenhando linhas no papel segundo o impulso. Você deve caminhar pela cidade e desenhar linhas no papel que correspondam ao que você vê. Se, vendo a cidade vagamente, você pensa que pode deslocar uma linha um pouco para a direita, um pouco para a esquerda, de acordo com seu capricho, este é o mesmo erro.

Na Ciência, quando se trata de decidir quais hipóteses testar, a moralidade da Ciência lhe dá liberdade pessoal para acreditar no que quiser, desde que não seja descartado por experimento e desde que você se mova para testar sua hipótese. A Ciência não tentaria dar um veredicto oficial sobre a melhor hipótese a ser testada, antes do experimento. Isso é deixado para a consciência do cientista individual.

Onde existe evidência experimental definitiva, a Ciência diz para você curvar seu pescoço teimoso e aceitá-la. Caso contrário, a Ciência deixa por sua conta. A Ciência lhe dá espaço para vagar nos limites da evidência experimental, de acordo com seus caprichos.

E isso não se concilia facilmente com a noção Bayesiana de uma estimativa de probabilidade exatamente correta, sem flexibilidade ou espaço para caprichos, que existe antes e depois do experimento. O Bayesianismo não combina bem com a razão antiga e tradicional para a Ciência - a desconfiança de grandes esquemas, a presunção de que as pessoas não são racionais o suficiente para acertar as coisas sem evidências experimentais definitivas e inconfundíveis. Se todos fôssemos Bayesianos perfeitos, não precisaríamos de um processo social de ciência.

Apesar disso, na época em que percebi meu grande erro, eu também estava estudando Kahneman, Tversky e Jaynes. Estava aprendendo um novo Caminho, mais rigoroso que a Ciência. Um Caminho que poderia criticar minha tolice, de uma forma que a Ciência nunca poderia. Um Caminho que poderia ter me dito o que a Ciência nunca teria dito com antecedência: “Você escolheu a hipótese errada para testar, idiota.”

Mas o Caminho de Bayes também é muito mais difícil de usar do que a Ciência. Ele coloca uma tremenda pressão sobre sua habilidade de ouvir pequenas notas falsas, onde a Ciência apenas exige que você perceba uma bigorna caindo em sua cabeça.

Na Ciência, você pode cometer um ou dois erros, e outro experimento virá e o corrigirá; na pior das hipóteses, você perde algumas décadas.

Mas se você tentar usar Bayes mesmo qualitativamente - se você tentar fazer a coisa que a Ciência não confia que você faça, e raciocinar racionalmente na ausência de evidências esmagadoras - é como matemática, pois um único erro em cem passos pode te levar a qualquer lugar. Exige leveza, uniformidade, precisão, perfeccionismo.

Há uma boa razão pela qual a Ciência não confia nos cientistas para fazer esse tipo de coisa e pede mais provas experimentais, mesmo depois que alguém afirma ter descoberto a resposta certa com base em dicas e lógica.

Mas se você prefere não perder dez anos tentando provar a teoria errada, precisará ensaiar o problema muito mais difícil: ouvir evidências que não gritam em seu ouvido.

Mesmo que você não consiga encontrar os priori para um problema no Manual de Química e Física - mesmo que não haja uma Fonte Autoritária lhe dizendo quais são os priori - isso não significa que você tenha uma escolha livre e pessoal de fazer os priori do jeito que você quiser. Isso significa que você tem um novo problema de adivinhação que você deve realizar da melhor maneira possível.

Se a mente, como um motor cognitivo, pudesse gerar estimativas corretas manipulando priori de acordo com caprichos, você poderia saber as coisas sem olhar para elas, ou até mesmo alterá-las sem as

tocar. Mas a mente não é mágica. A estimativa de probabilidade racional não tem espaço para nenhuma decisão baseada em capricho, mesmo quando parece que você não conhece os priori.

Da mesma forma, se a resposta Bayesiana é difícil de calcular, isso não significa que Bayes seja inaplicável; significa que você não sabe qual é a resposta Bayesiana. A teoria da probabilidade Bayesiana não é uma caixa de ferramentas de métodos estatísticos; é a lei que governa qualquer ferramenta que você usa, quer você saiba ou não, quer você possa calculá-la ou não.

Quanto a usar métodos Bayesianos em espaços de hipóteses enormes e altamente gerais - como, "Aqui estão os dados de todos os experimentos de física já realizados; agora, qual seria uma boa Teoria de Tudo?" - Se você soubesse como fazer isso, na prática, você não seria um estatístico, seria um programador de Inteligência Artificial Geral. Mas isso não significa que os seres humanos, ao modelar o universo usando a inteligência humana, estão violando as leis da física/Bayesianismo ao gerar palpites corretos sem evidência.

Nick Tarleton comenta:

> O problema é encorajar um padrão epistêmico privado tão frouxo quanto o social.

O que define o problema que eu estava tentando indicar muito melhor do que eu defini.

248 - Os cientistas já sabem disso?

poke alega:

> A capacidade de formular hipóteses relevantes é uma habilidade crucial, e ela consome grande parte do tempo de um cientista em desenvolvimento. Tal habilidade pode não estar explícita na descrição tradicional da ciência, mas isso não significa que não esteja embutida na real instituição social da ciência, que gera conhecimento concreto no mundo real. O problema reside na descrição e não na ciência em si.

Embora eu tenha chamado meu eu mais jovem de 'estúpido', isso é uma metáfora. Seria mais preciso dizer que eu gerenciava minha alta inteligência de maneira desajeitada. O jovem Eliezer aos 18 anos não tinha o hábito de cometer erros óbvios, mas o seu 'óbvio' diferia do meu.

Não, não passei pelo ensino tradicional. Porém, ao observar os erros cometidos por Eliezer aos 18 anos, percebo que muitos cientistas contemporâneos também cometem os mesmos equívocos. Não vejo indícios de que tenham sido mais bem instruídos do que eu.

Sir Roger Penrose, um físico de renome, ainda sustenta a ideia de que a consciência é causada pela gravidade quântica. É possível que nunca o tenham alertado sobre respostas misteriosas para questões igualmente misteriosas. Apenas lhe disseram que suas hipóteses precisavam ser testáveis e ter implicações empíricas, da mesma forma que ocorreu com Eliezer aos 18 anos.

'A consciência é causada pela gravidade quântica' tem implicações testáveis: implica que seria possível examinar os neurônios e encontrar uma superposição quântica coerente cujo colapso contribui para o processamento de informações, e que nunca seria viável reproduzir o comportamento de um neurônio em termos de entrada e saída por meio de uma simulação microanatômica computável...

… mas mesmo após afirmar 'A consciência é causada pela gravidade quântica', você não prevê como o cérebro pensa 'Penso, logo existo!' ou a natureza misteriosa do vermelho, embora acredite compreender a causa disso. Isso é um sinal de perigo notável, percebo eu agora, embora não seja o mesmo sinal de perigo do qual fui alertado. Duvido que Penrose tenha sido orientado nesse sentido por seu orientador de tese, assim como Niels Bohr não deve ter sido alertado ao desenvolver a Interpretação de Copenhague.

Tudo indica que o motivo pelo qual Eliezer aos 18 anos, Sir Roger Penrose e Niels Bohr não receberam um alerta é que não existe um alerta convencional.

Só generalizei o conceito de 'respostas misteriosas para perguntas misteriosas', com essas mesmas palavras, após elaborar uma análise bayesiana que diferencia explicações científicas técnicas, não técnicas e semi-técnicas. O resultado dessa análise pode ser resumido de forma não técnica por meio de quatro sinais de perigo:

- Em primeiro lugar, a explicação funciona como um limitador da curiosidade, em vez de um controlador da antecipação.
- Em segundo lugar, a hipótese não apresenta partes móveis; o fator misterioso não é um mecanismo complexo específico, mas uma substância ou força inexprimivelmente sólida.
- Terceiro, aqueles que oferecem a explicação, valorizam 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.

Em teoria, todos esses aspectos poderiam ter sido mencionados após o declínio do vitalismo. Assim como a teoria elementar da probabilidade poderia ter sido concebida por Arquimedes, ou os antigos gregos poderiam ter formulado a seleção natural. No entanto, ninguém me alertou explicitamente sobre nenhum desses quatro perigos, pelo menos não com esses termos, a não ser pelo aviso de que as hipóteses precisam ter implicações testáveis. E só concebi explicitamente os sinais de alerta ao tentar abordar o caso sob a perspectiva das distribuições de probabilidade, o que exigiu um certo grau de exagero.

Simplesmente não tenho motivos para acreditar que esses alertas sejam transmitidos no ensino científico, especialmente para a maioria dos cientistas. Isso inclui, entre outras coisas, conselhos sobre como lidar com situações de confusão, caos científico e desespero. Quando um cientista comum ou mentor médio teria oportunidade de usar esse tipo de técnica?

Acabamos de discutir o fiasco da física do mundo único. É evidente que ninguém os informou sobre a definição formal da Navalha de Occam, seja de forma explícita ou não.

Há um fenômeno conhecido em que grandes cientistas têm vários alunos excelentes. Isso pode se dever ao fato de os mentores transmitirem habilidades que não podem ser totalmente descritas. No entanto, isso não faz parte do padrão convencional da ciência. Se os grandes mentores não conseguiram expressar suas orientações em termos gerais e publicá-las, isso indica que esses conceitos não são compreendidos de maneira abrangente.

Raciocinar na ausência de evidências definitivas sem cometer erros imediatos é uma tarefa realmente árdua. Quando se está na escola, comete-se um erro, mas aprende-se cinquenta outros pontos corretos. No entanto, ao raciocinar sobre novos conhecimentos na ausência de orientação esmagadoramente convincente, pode-se cometer um erro e avançar cinquenta passos na direção errada.

Estou convencido de que muitos cientistas, ao saírem de suas especialidades, desligam suas mentes e aderem a crenças reconfortantes, sem compreenderem que as mentes são máquinas e cada crença confiável possui uma história causal. Duvido que lhes tenham dito que existe uma probabilidade racional precisa dada um certo estado de evidência, sem espaço para caprichos, mesmo que não seja possível calcular a resposta e não haja um comando oficial sobre o que acreditar.

É improvável que os cientistas que a mídia convida para especular sobre o futuro, criando detalhadas imagens do mundo em 2050, tenham sido alertados sobre a falácia da conjunção. Ou como a heurística da representatividade pode fazer com que histórias mais detalhadas pareçam mais plausíveis, embora cada detalhe adicional reduza a probabilidade. A ideia de que cada novo detalhe necessita de suporte independente e de não se poder inventar histórias complexas que pareçam idênticas às aprendidas em aulas de ciências ou história é crucial para o pensamento preciso na ausência de evidências definitivas. No entanto, como esse conceito se encaixaria no ensino científico convencional? O viés cognitivo só foi descoberto há algumas décadas e só se popularizou recentemente.

Além disso, há conceitos envolvendo noções como ‘emergência’ e ‘complexidade’, que são definidas de maneira vaga o suficiente para permitir uma variedade de interpretações. Existem subáreas inteiras da academia construídas em torno do tipo de erro que Eliezer aos 18 anos costumava cometer! (Embora eu nunca tenha me deixado levar por essa ideia de ‘emergência’).

Em certas ocasiões, costumo dizer que o objetivo da ciência consiste em acumular uma montanha tão grande de evidências que nem mesmo os próprios cientistas possam ignorá-la. Essa, de fato, é a característica que distingue um cientista, já que uma pessoa não-científica ignoraria essas evidências de qualquer forma.

Se pode existir alguma quantidade de evidência tão esmagadora que as pessoas finalmente desistam, abandonem teorias antigas e nunca mais as mencionem. Isso é suficiente para permitir que a catraca da Ciência gire para frente ao longo do tempo e crie uma civilização tecnológica. Isso contrasta com a religião.

Os livros de Carl Sagan, Martin Gardner e outras correntes do pensamento racional visam fazer essa

distinção: transformar alguém de um não-cientista em um potencial cientista e protegê-lo de crenças que tenham sido refutadas por evidências experimentais.

Que treinamento adicional recebem os cientistas profissionais? Algumas aulas em estatísticas frequentistas, ensinando como calcular a significância estatística, e treinamento em técnicas padronizadas que lhes permitem produzir trabalhos em um paradigma solidamente estabelecido.

Caso a Ciência exigisse mais do que isso do cientista médio, duvido que seria viável. Já enfrentamos problemas suficientes com pessoas que ingressam no campo sem as qualificações básicas.

Nick Tarleton resumiu bem o problema resultante, talvez até melhor do que eu: se você propuser uma hipótese aparentemente bizarra que ainda não foi refutada pelas evidências e tentar testá-la experimentalmente, a comunidade científica não o rotulará como uma pessoa ruim. A Ciência não confia em seus anciãos para determinar quais hipóteses 'não valem a pena testar'. No entanto, esse é um padrão social cuidadosamente negligente e, ao tentar aplicá-lo como um padrão de racionalidade epistêmica individual, você pode acabar acreditando demais. Voltando à analogia com o libertarianismo pragmático baseado na desconfiança, é adiferença entre 'Cigarros não deveriam ser ilegais' e 'Vá fumar um Marlboro'.

Você se lembra de alguém ter alertado especificamente sobre esse erro, em tantas palavras? Então, por que as pessoas não o cometeriam? Quantas pessoas vão espontaneamente se esforçar um pouco mais e ser ainda mais rigorosas consigo mesmas? Algumas, mas não muitas.

Muitos cientistas acreditam em várias coisas surpreendentes fora do ambiente de laboratório, caso possam se convencer de que tais crenças ainda não foram definitivamente refutadas ou que possam evitar fazê-las. Há alguma palestra padrão que os alunos de pós-graduação recebem na qual veem a tolice desse comportamento e perguntam: 'Eles perderam a aula naquele dia?' Na medida do que posso observar, não.

Talvez, se você tiver a sorte de ter um mentor renomado, ele lhe contará segredos pessoais raros, como ". Pergunte a si mesmo quais são os problemas importantes em sua área e trabalhe em um deles, em vez de optar por tarefas mais fáceis e triviais" ou "Seja mais cuidadoso do que o exigido pelos editores de revistas; procure novas maneiras de impedir que suas expectativas influenciem o experimento, mesmo que não seja a prática comum.

No entanto, não creio que exista uma ampla tradição científica de raciocínio epistêmico preciso em face de evidências escassas. Metade de todos os cientistas ainda acredita em Deus. As habilidades mais desafiadoras não fazem parte do padrão!

249 - Nenhuma defesa segura, nem mesmo a Ciência

Não costumo perguntar aos meus amigos sobre suas infâncias — falta-me curiosidade social — e por isso não sei o quanto isso é realmente uma tendência:

Das pessoas que conheço que estão se aprimorando como racionalistas, aquelas que voluntariamente compartilham informações sobre suas infâncias, há uma surpreendente propensão a ouvir coisas como: "Minha família se juntou a um culto e eu tive que escapar", ou "Um dos meus pais era clinicamente insano e eu tive que aprender a filtrar a realidade da loucura deles".

Minha própria experiência de crescer em uma família judia ortodoxa parece branda em comparação… mas conseguiu o mesmo resultado: quebrou minha confiança emocional fundamental na sanidade das pessoas ao meu redor.

Até que essa confiança emocional fundamental seja quebrada, você não começa a crescer como racionalista. Tenho dificuldade em explicar por que isso é assim. Talvez quaisquer habilidades incomuns que você adquira — qualquer coisa que o torne excepcionalmente racional — exija que você zigue quando outras pessoas zagam. Talvez isso seja assustador demais, se o mundo ainda parece um lugar são para você.

Ou talvez você não se incomode em fazer o trabalho árduo de ser extra bônus são, se a normalidade não o assusta.

Sei que muitos aspirantes a racionalistas parecem encontrar obstáculos em torno de coisas como criônica ou muitos mundos. Não que eles não vejam a lógica; eles veem a lógica e se perguntam: "Isso pode realmente ser verdade, quando parece tão óbvio agora, e ainda assim nenhuma das pessoas ao meu redor acredita nisso?"

Sim. Bem-vindo à Terra onde o etanol é feito de milho e ambientalistas se opõem à energia nuclear. Sinto muito.

(Veja também: Contraculturalismo Cultista. Se você acabar em um estado mental de buscar nervosamente reasseguramento, isso nunca é uma coisa boa — mesmo que seja porque você está prestes a acreditar em algo que soa lógico, mas poderia fazer outras pessoas olharem para você estranhamente.)

Pessoas que tiveram sua confiança quebrada na sanidade das pessoas ao seu redor parecem conseguir avaliar ideias estranhas por seus méritos, sem se sentirem nervosas com sua estranheza. A cola que as prende ao seu lugar atual se dissolveu, e elas podem caminhar em alguma direção, esperançosamente para frente.

Dissidência solitária, eu chamei isso. A verdadeira dissidência não se parece com ir à escola vestindo preto; se parece com ir à escola vestindo uma fantasia de palhaço.

É isso que é necessário para ser a voz solitária que diz: "Se você realmente acha que sabe quem vai ganhar a eleição, por que não está pegando o dinheiro grátis no mercado de previsões *Intrade*?" enquanto todas as pessoas ao seu redor estão pensando: "É bom ser um indivíduo e formar suas próprias opiniões, os comerciais de sapatos me disseram isso."

Talvez em algum outro mundo, algum ramo alternativo de Everett com uma população humana mais

sã, as coisas seriam diferentes... mas neste mundo, nunca vi ninguém começar a crescer como racionalista até que fizesse uma profunda ruptura emocional com a sabedoria de seu grupo.

Talvez em outro mundo, as coisas fossem diferentes. E talvez não. Não tenho certeza se os seres humanos realisticamente podem confiar e pensar ao mesmo tempo.

Era uma vez, havia algo em que eu confiava.

Eliezer18 confiava na Ciência.

Eliezer18 reconhecia obedientemente que o processo social da ciência era falho. Eliezer18 reconhecia obedientemente que o meio acadêmico era lento, alocava mal os recursos, tinha favoritismos e maltratava seus preciosos hereges.

Essa é a coisa conveniente sobre reconhecer falhas em pessoas que não conseguiram viver de acordo com seu ideal; você não precisa questionar o próprio ideal.

Mas quem poderia ser tolo o suficiente para questionar: "O método experimental decidirá qual hipótese vence"?

Parte do que enganou Eliezer18 foi um problema geral que ele tinha, uma aversão a ideias que se assemelhavam a coisas que idiotas haviam dito. Eliezer18 viu muitas pessoas questionando os ideais da Própria Ciência, e sem exceção, elas estavam todas do Lado das Trevas. Pessoas que questionavam o ideal da Ciência invariavelmente estavam tentando vender óleo de cobra, ou tentando proteger da crítica sua forma favorita de estupidez, ou tentando disfarçar sua resignação pessoal como uma aceitação Profundamente Sábia da futilidade.

Se houvesse qualquer outro ideal que tivesse alguns séculos de idade, o jovem Eliezer teria olhado para ele e dito: "Eu me pergunto se isso é realmente certo, e se há uma maneira de fazer melhor." Mas não o ideal da Ciência. A Ciência era a ideia mestra, a ideia que permitia que você mudasse ideias. Você poderia questioná-la, mas você deveria questioná-la e então aceitá-la, não realmente dizer: "Espere! Isso está errado!"

Assim, quando uma vez tive uma ideia estúpida, pensei estar me comportando virtuosamente se me certificasse de que havia uma Previsão Inédita, e professasse que desejava testar minha ideia experimentalmente. Pensei que tinha feito tudo o que era obrigado a fazer.

Então pensei estar seguro — não seguro de qualquer ameaça externa em particular, mas seguro em um nível mais profundo, como uma criança que confia em seus pais e obedeceu a todas as regras dos pais.

Há muito tempo eu tinha perdido a confiança na sanidade da minha família ou dos meus professores na escola. E as outras crianças não eram inteligentes o suficiente para competir com as conversas que eu podia ter com os livros. Mas eu confiava nos livros, entende. Eu confiava que se eu fizesse o que Richard Feynman me disse para fazer, eu estaria seguro. Nunca pensei essas palavras em voz alta, mas era assim que eu me sentia.

Quando Eliezer23 percebeu exatamente o quão estúpida a teoria estúpida tinha sido — e que a Racionalidade Tradicional não o havia salvado dela — e que a Ciência teria ficado perfeitamente bem com ele desperdiçando dez anos testando a ideia estúpida, desde que depois ele admitisse estar errado…

… bem, não direi que foi uma enorme convulsão emocional. Eu realmente não entro nesse tipo de drama. Simplesmente tornou-se óbvio que fui estúpido.

Essa é a confiança que estou tentando quebrar em você. Você não está seguro. Nunca.

Nem mesmo a Ciência pode salvá-lo. Os ideais da Ciência nasceram há séculos, em uma época em que ninguém sabia nada sobre teoria da probabilidade ou vieses cognitivos. A Ciência exige muito pouco de você, ela abençoa suas boas intenções com muita facilidade, não é rigorosa o suficiente, só faz aquelas injun-

ções que um cientista médio pode seguir, aceita a lentidão como um fato da vida.

Então não pense que se você apenas seguir as regras da Ciência, isso torna seu raciocínio defensável.

Não há nenhum procedimento conhecido que você possa seguir que torne seu raciocínio defensável.

Não há nenhum conjunto conhecido de injunções que você possa satisfazer e saber que não terá sido um tolo.

Não há nenhuma moralidade-de-raciocínio conhecida que você possa fazer o seu melhor para obedecer e saber que está, assim, protegido de críticas.

Não, nem mesmo se você se voltar para a Bayescracia. É muito mais difícil de usar e você nunca terá certeza de que está fazendo certo.

A disciplina da Bayescracia é muito mais jovem que a disciplina da Ciência. Você não encontrará livros didáticos, nem mentores idosos, nem histórias escritas de sucesso e fracasso, nem regras rígidas estabelecidas. Você terá que estudar vieses cognitivos, teoria da probabilidade, psicologia evolutiva, psicologia social e outras ciências cognitivas, e Inteligência Artificial — e pensar por si mesmo como aplicar todo esse conhecimento ao caso de corrigir a si, já que isso ainda não está nos livros didáticos.

Você não sabe o que sua própria mente está realmente fazendo. Eles encontram um novo viés cognitivo toda semana e você nunca tem certeza se o corrigiu ou o corrigiu em excesso.

A matemática formal é impossível de aplicar. Ela não se decompõe tão facilmente quanto João Q. Incrédulo pensa, mas você nunca tem realmente certeza de onde vêm os fundamentos. Você não sabe por que o universo é simples o suficiente para entender, ou por que qualquer prior funciona para ele. Você não sabe quais são seus próprios priori, muito menos se são bons.

Um dos problemas com a Ciência é que ela é vaga demais para realmente assustá-lo. “As ideias devem ser testadas por experimento.” Como você pode errar com isso?

Por outro lado, se você tem alguma matemática da teoria da probabilidade disposta na sua frente, e pior, você sabe que não pode realmente usá-la, então fica claro que você está tentando fazer algo difícil, e que você pode muito bem estar fazendo errado.

Então você não pode confiar.

E tudo isso que eu disse não será suficiente para quebrar sua confiança. Isso não acontecerá até que você entre em seu primeiro desastre real por seguir As Regras, não por quebrá-las.

Eliezer18 já tinha a noção de que era permitido questionar a Ciência. Ora, é claro que o método científico não era imune ao questionamento! Pois não somos todos bons racionalistas? Não temos permissão para questionar tudo?

Era a noção de que você poderia realmente na realidade seguir a Ciência e falhar miseravelmente que Eliezer18 não acreditava emocionalmente que fosse possível.

Oh, é claro que ele dizia ser possível. Eliezer18 reconhecia obedientemente a possibilidade de erro, dizendo: “Eu poderia estar errado, mas…”

Mas ele não achava que o fracasso pudesse acontecer, sabe, na realidade. Você deveria procurar falhas, não realmente encontrá-las.

E essa diferença emocional é uma coisa terrivelmente difícil de realizar em palavras, e temo que não haja maneira de eu realmente adverti-lo.

Sua confiança não se quebrará até que você aplique tudo o que aprendeu aqui e em outros livros, e leve isso tão longe quanto puder, e descubra que isso também falha com você — que você ainda foi um tolo, e

ninguém o advertiu contra isso — que todas as partes mais importantes foram deixadas de fora da orientação que você recebeu — que alguns dos ideais mais preciosos que você seguiu o guiaram na direção errada —

— e se você ainda tiver algo para proteger, de modo que deva continuar, e não possa renunciar e reconhecer sabiamente as limitações da racionalidade —

— então você estará pronto para começar sua jornada como racionalista. Para assumir a responsabilidade sozinho, viver sem defesas confiáveis e forjar uma Arte superior àquela que uma vez lhe foi ensinada.

Ninguém começa a buscar verdadeiramente o Caminho até que seus pais tenham falhado com eles, seus deuses estejam mortos e suas ferramentas tenham se estilhaçado em suas mãos.

Post Scriptum: Ao revisar um rascunho deste ensaio, descobri uma falha bastante imperdoável no raciocínio, que realmente afeta uma das conclusões tiradas. Estou deixando-a. Apenas caso você pensasse que seguir meu conselho o tornava seguro; ou que você deveria procurar falhas, mas não encontrar nenhuma.

E, é claro, se você procurar demais por uma falha, e encontrar uma falha que não é uma falha real, e se apegar a ela para se reassegurar de quão crítico você é, você só ficará pior do que antes…

É viver com a incerteza — saber em um nível visceral que existem falhas, elas são sérias e você não as encontrou — que é a coisa difícil.

250 - Mudando a definição de ciência

A *New Scientist* discute mudar a definição de ciência, disponível aqui: [1]

> Outros acreditam que tal crítica se baseia em um mal-entendido. "Algumas pessoas dizem que o conceito de multiverso não é falsificável porque é inobservável - mas isso é uma falácia", diz o cosmólogo Max Tegmark, do Instituto de Tecnologia de Massachusetts. Ele argumenta que o multiverso é uma consequência natural de teorias eminentemente falsificáveis, como a teoria quântica e a Relatividade Geral. Como tal, a teoria do multiverso se sustenta ou falha segundo o quão bem essas outras teorias resistem aos testes observacionais.
>
> [...]
>
> Então, se a simplicidade da falsificação é enganosa, o que os cientistas deveriam fazer em vez disso? Howson acredita que é hora de abandonar a noção de Popper de capturar o processo científico usando lógica dedutiva. Em vez disso, o foco deveria ser refletir o que os cientistas realmente fazem: reunir o peso da evidência para teorias rivais e avaliar sua plausibilidade relativa.
>
> Howson é um dos principais defensores de uma visão alternativa da ciência, baseada não na lógica simplista verdadeiro/falso, mas no conceito muito mais sutil de graus de crença. Em seu cerne está uma conexão fundamental entre o conceito subjetivo de crença e a matemática fria e dura da probabilidade[50].

Sou bem menos um iconoclasta solitário do que pareço. Talvez seja apenas o meu jeito de falar.

Os pontos de partida entre eu mesmo e a corrente dominante do "vamos-reformular-a-Ciência-como-Bayesianismo" são que:

1. Não estou na academia e posso me censurar muito menos quando se trata de dizer coisas "extremas" que outros podem muito bem já estar pensando.
2. Acho que **apenas ensinar teoria da probabilidade não será suficiente**. Teremos que sintetizar lições de várias ciências, como vieses cognitivos e psicologia social, formando uma nova Arte da Bayescracia coerente, antes de realmente melhorarmos no mundo real em relação à ciência moderna. A ciência tolera erros; a Bayescracia não. O ganhador do Nobel Robert Aumann, que primeiro provou que Bayesianos com os mesmos priori não podem concordar em discordar, é um judeu ortodoxo praticante. A teoria da probabilidade sozinha não resolverá o problema, quando se trata de realmente ensinar cientistas. Este é meu principal ponto de partida, e não é nada que vi sugerido em outro lugar.
3. Acho que é possível fazer melhor no mundo real. No caso extremo, uma superinteligência Bayesiana

50 NT. Texto original em inglês. *Others believe such criticism is based on a misunderstanding. "Some people say that the multiverse concept isn't falsifiable because it's unobservable—but that's a fallacy," says cosmologist Max Tegmark of the Massachusetts Institute of Technology. He argues that the multiverse is a natural consequence of such eminently falsifiable theories as quantum theory and General Relativity. As such, the multiverse theory stands or fails according to how well these other theories stand up to observational tests. [. . .]*
So if the simplicity of falsification is misleading, what should scientists be doing instead? Howson believes it is time to ditch Popper's notion of capturing the scientific process using deductive logic. Instead, the focus should be on reflecting what scientists actually do: gathering the weight of evidence for rival theories and assessing their relative plausibility. Howson is a leading advocate for an alternative view of science based not on simplistic true/false logic, but on the far more subtle concept of degrees of belief. At its heart is a fundamental connection between the subjective concept of belief and the cold, hard mathematics of probability.

poderia usar muito menos informação sensorial do que um cientista humano para chegar a conclusões corretas. Na primeira vez que você vê uma maçã cair, você observa que a posição varia com o quadrado do tempo, inventa o cálculo, generaliza as Leis de Newton... e vê que as Leis de Newton envolvem ação à distância, procura explicações alternativas com maior localidade, inventa a covariância relativística em torno de um limite de velocidade hipotético e considera que a Relatividade Geral pode valer a pena ser testada.

Os humanos não processam evidências eficientemente - nossas mentes são tão barulhentas que exigem ordens de magnitude a mais de evidências extras para nos colocar de volta nos trilhos depois que descarrilhamos. Nosso coletivo, o meio acadêmico, é ainda mais lento.

Referências

[1] Robert Matthews, “Do We Need to Change the Definition of Science?,” New Scientist (May 2008).

251 - Mais rápido que a Ciência

Às vezes digo que o método da ciência é acumular uma montanha tão enorme de evidências que nem mesmo os cientistas podem ignorá-la; e que essa é a característica distintiva de um cientista. (Um não-cientista a ignorará de qualquer maneira.)

Max Planck era ainda menos otimista: [1]

> Uma nova verdade científica não triunfa ao convencer seus oponentes e fazê-los ver a luz, mas sim porque seus oponentes eventualmente morrem, e uma nova geração cresce que está familiarizada com ela[51].

Acho essa noção muito divertida, porque implica que o poder da ciência para distinguir verdade de falsidade depende, em última análise, do bom gosto dos estudantes de pós-graduação.

O aumento gradual na aceitação de muitos mundos na física acadêmica sugere haver físicos que só aceitarão uma nova ideia com alguma combinação de justificação epistêmica e um grupo acadêmico suficientemente grande no qual possam se sentir confortáveis. À medida que mais físicos aceitam, o grupo cresce e, assim, mais pessoas ultrapassam seus limites individuais para conversão—com a justificação epistêmica permanecendo essencialmente a mesma.

Mas a Ciência ainda chega lá eventualmente, e isso é suficiente para que o mecanismo da Ciência avance e crie uma civilização tecnológica.

Os cientistas podem ser movidos por preconceitos infundados, por intuições debilitadas, por comportamento de manada—a panóplia de falhas humanas. Cada vez que um cientista muda de crença por razões epistemologicamente injustificáveis, são necessárias mais evidências ou novos argumentos para cancelar o ruído.

O "colapso da função de onda" não tem justificação experimental, mas apela à intuição (debilitada) de um único mundo. Então pode ser necessário um argumento extra—digamos, que o colapso viola a Relatividade Especial—para iniciar a lenta desintegração acadêmica de uma ideia que nunca deveria ter sido atribuída uma probabilidade não negligenciável em primeiro lugar.

De uma perspectiva bayesiana, a ciência acadêmica humana na totalidade é um processador de evidências altamente ineficiente. Cada vez que um argumento injustificável muda a crença, é necessário um argumento justificável extra para mudá-la de volta. O processo social da ciência se apoia em evidências extras para superar o ruído cognitivo.

Uma maneira mais caridosa de dizer isso é que os cientistas adotarão posições que são teoricamente insuficientemente extremas, em comparação com as posições ideais que os cientistas adotariam, se fossem IAs bayesianas e confiassem em si mesmos para raciocinar claramente.

Mas não seja muito caridoso. O ruído de que estamos falando não é apenas de erros inocentes. Em muitos campos, os debates se arrastam por décadas depois que deveriam ter sido resolvidos. E não porque os

51 NT. Texto original em inglês. *A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it.*

cientistas de ambos os lados se recusam a confiar em si mesmos e concordam que deveriam procurar evidências adicionais. Mas, porque um lado continua lançando objeções ridículas e exigindo mais e mais evidências, a partir de uma posição entrincheirada de poder acadêmico, muito após ficar claro de onde sopram os ventos das evidências. (Estou pensando aqui nos debates em torno da invenção da psicologia evolutiva, não sobre muitos mundos.)

É possível que humanos individuais ou grupos processem evidências de maneira mais eficiente—cheguem a conclusões corretas mais rapidamente—do que a ciência acadêmica humana na totalidade?

"As ideias são testadas por experimentos. Esse é o cerne da ciência." E isso deve ser verdade, porque se você não pode confiar no Feynman Zumbi, em quem você pode confiar?

Ainda assim, de onde vêm as ideias?

Você pode ser tentado a responder, "Elas vêm dos cientistas. Tem outra pergunta?" Na Ciência você não deve se importar de onde vêm as hipóteses—apenas se elas passam ou falham experimentalmente.

Ok, mas se você remover todas as novas ideias, o processo científico como um todo para de funcionar porque não tem hipóteses alternativas para testar. Então, inventar novas ideias não é uma parte dispensável do processo.

Agora coloque seus óculos bayesianos de volta. Como descrito em "A Arrogância de Einstein", há consultas que não são binárias—onde a resposta não é "Sim" ou "Não", mas retirada de um espaço maior de estruturas, por exemplo, o espaço de equações. Em tais casos, são necessárias muito mais evidências bayesianas para promover uma hipótese à sua atenção do que para confirmar a hipótese.

Se você está trabalhando no espaço de todas as equações que podem ser especificadas em 32 bits ou menos, está trabalhando em um espaço de 4 bilhões de equações. São necessárias muito mais evidências bayesianas para elevar uma dessas hipóteses ao nível de 10% de probabilidade, do que para elevar ainda mais a hipótese de 10% para 90% de probabilidade.

Quando o espaço de ideias é grande, criar ideias dignas de teste envolve muito mais trabalho—no sentido termodinâmico bayesiano de "trabalho"—do que simplesmente obter um resultado experimental com $p < 0{,}0001$ para a nova hipótese sobre a hipótese antiga.

Se isso não parecer óbvio à primeira vista, pare aqui e reveja "A Arrogância de Einstein".

O processo científico sempre confiou nos cientistas para criar hipóteses a serem testadas, por meio de algum processo não especificado pela ciência. Suponha que você tenha desenvolvido uma maneira de gerar hipóteses que fosse completamente louca—digamos, bombear um tabuleiro Ouija controlado por robô com os dígitos de pi—e as sugestões resultantes continuassem sendo verificadas experimentalmente. A essência pura e ideal da Ciência não perderia o ritmo. A essência pura e ideal de Bayes se incendiaria e morreria.

(Comparado à ciência, Bayes é falsificado por mais dos possíveis resultados.)

Isso não significa que o processo de decidir quais ideias testar seja sem importância para a ciência. Significa que a ciência não o especifica.

Na prática, o tabuleiro Ouija controlado por robô não funciona. Na prática, há algumas consultas científicas com um espaço de resposta grande o suficiente que, escolhendo modelos aleatoriamente para testar, levaria zilhões de anos para encontrar um modelo que fizesse boas previsões—como fazer macacos digitarem Shakespeare.

Na fronteira da ciência—a fronteira entre a ignorância e o conhecimento, onde a ciência avança—o processo depende de pelo menos alguns cientistas individuais (ou grupos de trabalho) verem coisas que ainda não foram confirmadas pela ciência. É assim que eles sabem quais hipóteses testar, antes do próprio teste.

Se você tirar seus óculos bayesianos, pode dizer: "Bem, eles não precisam saber, apenas precisam

adivinhar." Se você colocar seus óculos bayesianos de volta, perceberá que "adivinhar" com 10% de probabilidade requer quase tanto trabalho epistêmico realizado, nos bastidores, quanto "adivinhar" com 80% de probabilidade—pelo menos para grandes espaços de resposta.

O cientista pode não saber que realizou esse trabalho epistêmico com sucesso, antes do experimento; mas ele deve, de fato, tê-lo feito com sucesso! Caso contrário, ele nem sequer pensaria na hipótese correta. Pelo menos em grandes espaços de resposta.

Então o cientista faz a previsão inédita, realiza o experimento, publica o resultado, e agora a Ciência também sabe. Agora faz parte do conhecimento publicamente acessível da humanidade, que qualquer um pode verificar por si mesmo.

Entre esses momentos houve um intervalo no qual o cientista sabia racionalmente algo que o processo social público da ciência ainda não havia confirmado. E esse não é um intervalo trivial, embora possa ser curto; pois é onde reside a fronteira da ciência, a fronteira dos avanços.

Tudo isso é mais verdadeiro para a ciência não rotineira do que para a ciência rotineira, porque é uma noção de grandes espaços de resposta onde a resposta não é "Sim" ou "Não" ou retirada de um pequeno conjunto de alternativas óbvias. É muito mais fácil treinar pessoas para testar ideias do que ter boas ideias para testar.

Referências

[1] Max Planck, Scientific Autobiography and Other Papers (New York: Philosophical Library, 1949).

252 - A velocidade de Einstein

No ensaio anterior, argumentei que os Poderes Além da Ciência são, na verdade, uma parte padrão e necessária do processo social da ciência. Em particular, os cientistas devem recorrer aos seus poderes de racionalidade individual para decidir quais ideias testar. Isso ocorre antes do tipo de experimentos definitivos que a Ciência exige para abençoar uma ideia como confirmada. O ideal da Ciência não tenta especificar esse processo. Não supomos que alguma autoridade pública saiba como os cientistas individuais devem pensar. Mas isso não significa que o processo não seja importante.

Um exemplo facilmente compreensível e não perturbador:

Um cientista identifica uma forte regularidade matemática nos dados acumulados de experimentos anteriores. Mas a hipótese correspondente ainda não fez nem confirmou uma nova previsão experimental - o que seu campo acadêmico exige; este é um daqueles campos onde você pode realizar experimentos controlados sem muitos problemas. Assim, o cientista individual tem razões racionais e facilmente compreensíveis para acreditar (embora não com probabilidade 1) em algo que ainda não foi abençoado pela Ciência como conhecimento público da humanidade.

Notar uma regularidade em uma grande massa de dados experimentais não parece tão não científico. Você ainda é orientado por dados, certo?

Mas isso é porque eu deliberadamente escolhi um exemplo não perturbador. Quando Einstein inventou a Relatividade Geral, ele não tinha quase nenhum dado experimental para se basear, exceto a precessão do periélio de Mercúrio. E (até onde sei) Einstein não usou esses dados, exceto no final.

Einstein gerou a teoria da Relatividade Especial usando o Princípio de Mach, o qual é a versão dos físicos do Princípio Generalizado Anti-Zumb. Você começa dizendo: “Não me parece razoável que você pudesse dizer, em uma sala fechada, quão rápido você e a sala estavam indo. Como esse número não deveria ser observável, ele não deveria existir em nenhum sentido significativo.” Você então observa que as Equações de Maxwell invocam uma velocidade de propagação aparentemente absoluta, c, comumente referida como “a velocidade da luz” (embora as equações quânticas mostrem que é a velocidade de propagação de todas as ondas fundamentais). Então você reformula sua física de tal maneira que a velocidade absoluta de um único objeto não exista mais significativamente, e apenas velocidades relativas existam. Estou pulando muita coisa aqui, obviamente, mas há muitas excelentes introduções à relatividade. Não é como a horrível situação na física quântica.

Einstein, tendo conseguido se livrar da noção de sua velocidade absoluta numa sala fechada, então se propôs a se livrar da noção de sua aceleração absoluta numa sala fechada. Parecia a Einstein que não deveria haver uma maneira de diferenciar, em uma sala fechada, entre a sala acelerando para o norte enquanto o resto do universo ficava parado, versus o resto do universo acelerando para o sul enquanto a sala ficava parada. Se o resto do universo acelerasse, produziria ondas gravitacionais que te acelerariam. Matéria em movimento, então, deveria produzir ondas gravitacionais.

E, porque a massa inercial e a massa gravitacional eram sempre exatamente equivalentes - ao contrário da situação no eletromagnetismo, onde um elétron e um múon podem ter massas diferentes, mas a mesma carga elétrica - a gravidade deveria se revelar como um tipo de inércia. A Terra deveria girar em torno do Sol em algo equivalente a uma “linha reta”. Isso requer que o espaço-tempo nas proximidades do Sol seja

curvo. Assim, se você desenhasse um gráfico da órbita da Terra ao redor do Sol, a linha no papel gráfico 4D seria localmente plana. Então a massa inercial e gravitacional seriam necessariamente equivalentes, não apenas coincidentemente equivalentes.

(Se isso não fez sentido para você, há boas introduções à Relatividade Geral disponíveis também.)

E, é claro, a nova teoria tinha que obedecer à Relatividade Especial, conservar energia, conservar momento, e assim por diante.

Einstein passou vários anos compreendendo a matemática necessária para descrever métricas curvas do espaço-tempo. Então ele escreveu a teoria mais simples que tinha as propriedades que Einstein achava que deveria ter. Isso incluía propriedades que ninguém jamais havia observado, mas que Einstein achava que se encaixavam bem com o caráter de outras leis físicas. Então Einstein fez alguns cálculos e obteve a precessão anteriormente inexplicada de Mercúrio de volta.

Quão impressionante foi isso?

Bem, coloquemos desta forma. Em alguma pequena fração de Terras alternativas procedendo de 1800 - talvez até uma fração considerável - pareceria plausível que a física relativística pudesse ter procedido de maneira similar ao nosso grande fiasco com a física quântica.

Podemos imaginar que a "interpretação" original de Lorentz da contração de Lorentz, como uma distorção física causada pelo movimento em relação ao éter, prevaleceu. Podemos imaginar que vários fatores corretivos, eles próprios inexplicados, foram adicionados à mecânica gravitacional newtoniana para explicar a precessão de Mercúrio. Talvez atribuídos a estranhas distorções do éter, como na contração de Lorentz. Ao longo das décadas, mais fatores corretivos seriam adicionados para explicar outras observações astronômicas. Relógios atômicos suficientemente precisos, em aviões, revelariam que o tempo corria um pouco mais rápido do que o esperado em altitudes mais altas (o tempo corre mais devagar em campos gravitacionais mais intensos, mas eles não saberiam disso) e mais "fatores etéreos" corretivos seriam inventados.

Até que, finalmente, os muitos diferentes "fatores corretivos" determinados empiricamente fossem unificados nas equações simples da Relatividade Geral.

E as pessoas nessa Terra alternativa diriam: "A equação final era simples, mas não havia como você saber chegar a essa resposta apenas com a precessão do periélio de Mercúrio. São necessários muitos, muitos experimentos adicionais. Você deve ter medido o tempo correndo mais devagar em um campo gravitacional mais forte; você deve ter medido a luz se curvando ao redor das estrelas. Só então você pode imaginar nossa teoria unificada da gravitação etérea. Não, nem mesmo uma superinteligência bayesiana perfeita poderia saber! - pois haveria muitas teorias ad-hoc consistentes apenas com a precessão do periélio."

Em nosso mundo, Einstein nem mesmo usou a precessão do periélio de Mercúrio, exceto para verificação de sua resposta produzida por outros meios. Einstein sentou-se em sua poltrona e pensou em como ele teria projetado o universo, para parecer do jeito que ele achava que um universo deveria parecer. Por exemplo, que você não deveria conseguir distinguir você mesmo acelerando em uma direção, do resto do universo acelerando na outra direção.

E Einstein executou toda a longa (de vários anos!) cadeia de raciocínio de poltrona, sem cometer erros que teriam exigido mais evidências experimentais para colocá-lo de volta no caminho certo.

Até mesmo Jeffreyssai ficaria relutantemente impressionado. Embora ele ainda tiraria um ponto ou dois de Einstein pela constante cosmológica. (Eu não tiro pontos de Einstein pela constante cosmológica porque mais tarde ela se revelou real. Tento evitar criticar pessoas em ocasiões em que elas estão certas.)

Qual seria a perspectiva teórica de probabilidade sobre o feito de Einstein?

Em vez de observar os planetas e inferir que leis poderiam cobrir sua gravitação, Einstein estava observando as outras leis da física e inferindo que nova lei poderia seguir o mesmo padrão. Einstein não estava encontrando uma equação que cobrisse o movimento dos corpos gravitacionais. Einstein estava encontrando

um caráter-de-lei-física que cobria equações previamente observadas, e que ele poderia usar para prever a próxima equação que seria observada.

Ninguém sabe de onde vem as leis da física, mas o sucesso de Albert Einstein com a Teoria da Relatividade Geral sugere que essas leis possuem um caráter comum robusto. Esse caráter permite a previsão da forma correta de uma lei ao observar outras leis relacionadas, sem necessariamente precisar observar os efeitos exatos de cada lei individualmente.

(Em um sentido geral, é claro, Einstein sabia por observação que as coisas caíam; mas ele não obteve a Relatividade Geral por inferência retrógrada do avanço exato do periélio de Mercúrio.)

Então, de uma perspectiva bayesiana, o que Einstein fez ainda é indução, e ainda é coberto pela noção de um prior simples (prior de Ocam) atualizado por novas evidências. É apenas o prior que era sobre os possíveis caracteres da lei física, e observar outras leis físicas permitiu que Einstein atualizasse seu modelo do caráter da lei física, que ele então usou para prever uma lei particular da gravitação.

Se você não tivesse o conceito de um "caráter de lei física", o que Einstein fez pareceria mágica - tirando o modelo correto de gravitação do nada entre todas as equações possíveis, com evidências vastamente insuficientes. Mas Einstein, ao olhar para outras leis, reduziu o espaço de possibilidades para a próxima lei. Ele aprendeu o alfabeto no qual a física era escrita, restrições para governar sua resposta. Não mágica, mas raciocínio em um nível mais alto, através de um domínio mais amplo, do que o que um raciocinador ingênuo poderia conceber como o "espaço modelo" de apenas esta única lei.

Então, do ponto de vista teórico de probabilidade, Einstein ainda era orientado por dados - ele apenas usou os dados que já tinha, de forma mais eficaz. Comparado a quaisquer Terras alternativas que exigissem grandes quantidades de dados adicionais de observações astronômicas e relógios em aviões para bater na cabeça deles com a Relatividade Geral.

Há numerosas lições que podemos derivar disso.

Uso Einstein como meu exemplo, mesmo que seja clichê, porque Einstein também era incomum no sentido de que admitia abertamente saber coisas que a Ciência não havia confirmado. Perguntado o que teria feito se a observação do eclipse solar de Eddington tivesse falhado em confirmar a Relatividade Geral, Einstein respondeu: "Então eu sentiria pena do bom Deus. A teoria está correta."

Segundo as noções prevalecentes de Ciência, isso é arrogância - você deve aceitar o veredicto do experimento e não se apegar às suas ideias pessoais.

Mas como concluí em "A Arrogância de Einstein", Einstein não se sai tão mal de uma perspectiva bayesiana. De uma perspectiva bayesiana, para sugerir a Relatividade Geral, para sequer pensar no que acabou sendo a resposta correta, Einstein deve ter tido evidências suficientes para identificar a resposta verdadeira no espaço-teoria. Seria necessário apenas um pouco mais de evidências para justificar (em um sentido bayesiano) estar quase certo da teoria. E era improvável que Einstein tivesse apenas exatamente evidências suficientes para trazer a hipótese completamente à sua atenção.

Qualquer acusação de arrogância teria que se centrar na questão: "Mas Einstein, como você sabia que tinha raciocinado corretamente?" - Ao que eu só posso dizer: não critique as pessoas quando elas acabam estando certas! Espere por uma ocasião em que elas estejam erradas! Caso contrário, você está perdendo a chance de ver quando alguém está pensando de forma mais inteligente que você - pois você os critica sempre que se afastam de um ritual preferido de cognição.

Ou considere a famosa troca entre Einstein e Niels Bohr sobre a teoria quântica - em um momento em que a então atual teoria quântica de mundo único parecia imensamente bem confirmada experimentalmente; um tempo em que, pelos padrões da Ciência, a atual teoria quântica (desequilibrada) havia simplesmente vencido.

EINSTEIN: "Deus não joga dados com o universo."

BOHR: “Einstein, não diga a Deus o que fazer.”

Você tem que admirar alguém que pode entrar em uma discussão com Deus e vencer.

Se você tirar seus óculos bayesianos e olhar para Einstein em termos do que ele fazia realmente o dia todo, então o cara estava sentado estudando matemática e pensando em como ele projetaria o universo, em vez de sair e olhar para as coisas para coletar mais dados. O que Einstein fez, com sucesso, é exatamente o tipo de feito intelectual elevado de puro intelecto que Aristóteles pensava que poderia fazer, mas não conseguia. Não de uma posição teórica de probabilidade, note bem, mas do ponto de vista do que eles faziam o dia todo.

A Ciência não confia nos cientistas para fazer isso, é por isso que a Relatividade Geral não foi abençoada como o conhecimento público da humanidade até após ter feito e verificado uma nova previsão experimental - tendo a ver com a curvatura da luz em um eclipse solar. (Mais tarde, descobriu-se que aquela medição particular não era precisa o suficiente para verificar de forma confiável, e havia favorecido a Relatividade Geral essencialmente por sorte.)

No entanto, só porque a Ciência não confia nos cientistas para fazer algo, não significa que seja impossível.

Mas uma palavra de cautela aqui: A razão pela qual os livros de história às vezes registram os nomes de cientistas que pensaram grandes pensamentos elevados não é que o pensamento elevado seja mais fácil ou mais confiável. É um viés de prioridade: Algum cientista que raciocinou com sucesso a partir da menor quantidade de evidência experimental chegou à verdade primeiro. Isso não pode ser uma questão de puro acaso: O espaço da teoria é muito grande, e Einstein ganhou várias vezes seguidas. Mas de todos os cientistas que tentaram desvendar um enigma, ou que teriam eventualmente tendo sucesso com evidências suficientes, a história nos transmite os nomes dos cientistas que conseguiram chegar lá primeiro. Tenha isso em mente quando estiver tentando derivar lições sobre como raciocinar prudentemente.

No dia a dia, você quer cada pedaço de evidência que puder obter. Não confie na sua capacidade de pensar elevadamente, a menos que a experimentação seja tão cara ou perigosa que você não tenha outra escolha.

Às vezes, porém, os experimentos são caros, e às vezes preferimos chegar lá primeiro... então você pode considerar treinar-se para raciocinar com evidências escassas, preferencialmente em casos onde depois descobrirá se estava certo ou errado. Tentar superar mercados de previsão de baixa capitalização pode ser um bom treino para isso? — embora isso seja apenas especulação.

Até agora, pelo menos, o raciocínio baseado em evidências escassas é algo que a ciência moderna não consegue treinar os cientistas modernos para fazer de forma alguma. O que talvez tenha algo a ver com, não sei, nem mesmo tentar?

Na verdade, retiro o que disse. O pensamento mais sensato que já vi em qualquer campo científico vem da psicologia evolutiva, possivelmente porque eles entendem o autoengano, mas talvez também porque frequentemente (1) têm que raciocinar com evidências escassas e (2) depois descobrem se estavam certos ou errados. Recomendo a todos os aspirantes a racionalistas que estudem psicologia evolutiva simplesmente para vislumbrar como é um raciocínio cuidadoso. Veja especialmente *The Psychological Foundations of Culture* (As bases psicológicas da cultura) de Tooby e Cosmides. [1]

Quanto à possibilidade de que apenas Einstein poderia fazer o que Einstein fez… que isso exigia superpoderes além do alcance dos mortais comuns… aqui esbarramos em alguns vieses que exigiriam um ensaio separado para analisar. Deixe-me colocar desta forma: É possível, talvez, que apenas um gênio pudesse ter feito o trabalho histórico real de Einstein. Mas gênios em potencial, em termos de inteligência bruta, são provavelmente muito mais comuns do que super realizadores históricos. Para colocar um número aleatório nisso, duvido que seja necessário mais do que um fator-g de um em um milhão para ser um gênio em potencial de classe mundial, o que implica pelo menos seis mil Einsteins em potencial circulando por aí hoje. E quanto a todos os outros, não vejo razão para não aspirarem a usar eficientemente as evidências que têm.

Mas minha moral final é que a fronteira onde o cientista individual racionalmente sabe algo que a Ciência ainda não confirmou nem sempre é uma questão inocentemente orientada por dados de identificar uma forte regularidade em uma montanha de experimentos. Às vezes, o cientista chega lá pensando grandes pensamentos elevados que a Ciência não confia que você pense.

Não direi "Não tente isso em casa". Direi "Não pense que isso é fácil". Não estamos discutindo, aqui, a vitória de opiniões casuais sobre cientistas profissionais. Estamos discutindo as ocasionais vitórias históricas de um tipo de esforço profissional sobre outro. Nunca se esqueça de todos os famosos casos históricos em que tentativas de raciocínio de poltrona fracassaram.

Referências

[1] Tooby and Cosmides, "The Psychological Foundations of Culture."

253 - Aquela mensagem alienígena

Imagine um mundo muito parecido com este, no qual, graças às tecnologias de seleção genética, o QI médio é 140 (na nossa escala). Einsteins em potencial são um em mil, não um em um milhão; e eles crescem em um sistema escolar adequado, se não para eles pessoalmente, pelo menos para crianças brilhantes. Cálculo é rotineiramente ensinado na sexta série. O próprio Albert Einstein ainda viveu e fez aproximadamente as mesmas descobertas, mas seu trabalho não parece mais excepcional. Vários físicos modernos de ponta fizeram avanços equivalentes e ainda estão por aí para conversar.

(Não, este não é o mundo em que Brennan vive.)

Um dia, as estrelas no céu noturno começam a mudar.

Algumas ficam mais brilhantes. Outras ficam mais fracas. A maioria permanece a mesma. Telescópios astronômicos capturam tudo, momento a momento. As estrelas que mudam, mudam sua luminosidade uma de cada vez, distintamente; a mudança de luminosidade ocorre no decorrer de um microssegundo, mas um segundo inteiro separa cada mudança.

Fica claro, desde o primeiro instante em que alguém percebe que mais de uma estrela está mudando, que o processo parece se concentrar particularmente na Terra. A chegada da luz dos eventos, em muitas estrelas espalhadas pela galáxia, foi precisamente cronometrada para a Terra em sua órbita. Logo, a confirmação vem de telescópios em órbita alta (eles têm esses) de que os milagres astronômicos não parecem tão sincronizados de fora da Terra. Somente os telescópios da Terra veem uma estrela mudando a cada segundo (1005 milissegundos, na verdade).

Quase toda a capacidade intelectual combinada da Terra se volta para a análise.

Rapidamente fica claro que as estrelas que aumentam em luminosidade aumentam em um fator de exatamente 256; aquelas que diminuem em luminosidade diminuem em um fator de exatamente 256. Não há padrão aparente nas coordenadas estelares. Isso deixa, simplesmente, um padrão de brilhante-fraco-brilhante-brilhante...

“Uma mensagem binária!” é o primeiro pensamento de todos.

Mas neste mundo existem pensadores cuidadosos, de grande prestígio também, e eles não têm tanta certeza. “Existem maneiras mais fáceis de enviar uma mensagem”, eles postam em seus blogs, “se você pode fazer as estrelas piscarem e se quiser se comunicar. Algo está acontecendo. Parece, prima facie, focar na Terra em particular. Chamar isso de ‘mensagem’ pressupõe muito mais sobre a causa por trás disso. Pode haver algum tipo de processo evolutivo entre, hum, coisas que podem fazer as estrelas piscarem, que acaba sendo sensível à inteligência de alguma forma... Sim, provavelmente há algo como ‘inteligência’ por trás disso, mas tente apreciar a ampla gama de possibilidades que isso realmente implica. Não sabemos se esta é uma mensagem, ou se foi enviada com o mesmo tipo de motivações que poderiam nos mover. Quer dizer, nós apenas sinalizaríamos usando uma lanterna grande, não bagunçaríamos uma galáxia inteira.”

Nesse momento, alguém começou a coletar os dados astronômicos e postá-los na Internet. Sugestões iniciais de que os dados podem ser prejudiciais foram... não ignoradas, mas também não obedecidas. Se algo tão poderoso quiser te machucar, você está praticamente morto (as pessoas raciocinam).

Vários grupos de pesquisa estão procurando por padrões nas coordenadas estelares - ou tempos de chegada fracionários das mudanças, em relação ao centro da Terra - ou durações exatas da mudança de luminosidade - ou qualquer pequena variação na mudança de magnitude - ou qualquer outro fato que possa ser conhecido sobre as estrelas antes de elas mudarem. Mas a maioria das pessoas está voltando sua atenção para o padrão de brilhantes e fracos.

Rapidamente fica claro que o padrão enviado é altamente redundante. Dos primeiros 16 bits, 12 são brilhantes e 4 são fracos. Os primeiros 32 bits recebidos se alinham com os segundos 32 bits recebidos, com apenas 7 dos 32 bits diferentes, e então os próximos 32 bits recebidos têm apenas 9 dos 32 bits diferentes do segundo (e 4 deles são bits que mudaram antes). Dos primeiros 96 bits, então, fica claro que esse padrão não é uma codificação ideal e compactada de nada. O pensamento óbvio é que a sequência pretende transmitir instruções para decodificar uma mensagem compactada a seguir...

"Mas", dizem os pensadores cuidadosos, "qualquer um que se importasse com eficiência, com poder suficiente para mexer com estrelas, talvez pudesse ter apenas nos sinalizado com uma lanterna grande e nos enviado um DVD?"

Também parece haver estrutura dentro dos grupos de 32 bits; alguns subgrupos de 8 bits ocorrem com maior frequência do que outros, e essa estrutura só aparece ao longo dos alinhamentos naturais (32 = 8 + 8 + 8 + 8).

Após as primeiras cinco horas a um bit por segundo, uma redundância adicional se torna clara: a mensagem começou a se repetir aproximadamente no 16.385º bit.

Dividindo a mensagem em grupos de 32, há 7 bits de diferença entre o 1º grupo e o 2º grupo, e 6 bits de diferença entre o 1º grupo e o 513º grupo.

"Uma imagem 2D!", todos pensam. "E os quatro grupos de 8 bits são cores; eles são tetracromatas!"

Mas logo fica claro que há uma assimetria horizontal/vertical: menos bits mudam, em média, entre (N; N + 1) versus (N; N + 512). O que você não esperaria se a mensagem fosse uma imagem 2D projetada em uma grade simétrica. Então você esperaria que a distância média bit a bit entre dois grupos de 32 bits fosse como a norma-2 da separação da grade: .

Também se forma um consenso geral de que uma certa codificação binária de 8 grupos em números inteiros entre -64 e 191 - não a codificação binária que nos parece óbvia, mas ainda altamente regular - minimiza a distância média entre células vizinhas. Isso continua a ser confirmado pelos bits recebidos.

Estatísticos, criptógrafos, físicos e cientistas da computação vão trabalhar. Há estrutura aqui; ela só precisa ser desvendada. Os mestres da causalidade procuram por independência condicional, triagem e vizinhanças de Markov, entre bits e grupos de bits. A chamada "cor" parece desempenhar um papel nas vizinhanças e na triagem, então não é apenas o equivalente da refletividade da superfície. As pessoas procuram por equações simples, autômatos celulares simples, árvores de decisão simples, que possam prever ou comprimir a mensagem. Físicos inventam teorias inteiramente novas da física que poderiam descrever universos projetados na grade - pois parece bastante plausível que uma mensagem como esta esteja sendo enviada de além da Matrix.

Após receber 32 * 512 * 256 = 4.194.304 bits, cerca de um mês e meio, as estrelas param de piscar.

O trabalho teórico continua. Físicos e criptógrafos arregaçam as mangas e começam a trabalhar seriamente. Eles resolveram problemas com muito menos dados do que isso. Físicos testaram teorias inteiras com pequenas diferenças de massa de partículas; criptógrafos desvendaram mensagens mais curtas deliberadamente obscurecidas.

Anos se passam.

Dois modelos dominantes sobreviveram, no meio acadêmico, no escrutínio do público e no escrutínio daqueles cientistas que antes faziam trabalhos semelhantes aos de Einstein. Há uma teoria de que a grade

é uma projeção de objetos em um espaço de 5 dimensões, com uma assimetria entre 3 e 2 das dimensões espaciais. Há também uma teoria de que a grade se destina a codificar um autômato celular - sem dúvida, a grade tem várias propriedades favoráveis para isso. Códigos foram criados que mostram comportamentos interessantes; mas até agora, executar os autômatos correspondentes nos maiores computadores disponíveis não conseguiu produzir nenhum resultado decodificável. A execução continua.

De vez em quando, um grupo de jovens estudantes especialmente brilhantes, que nunca viram a sequência binária detalhada, é selecionado. Esses alunos veem apenas as primeiras 32 linhas (de 512 colunas cada), para ver se podem formar novos modelos e quão bem esses novos modelos preveem as próximas 224 linhas. Tanto o modelo 3+2 dimensional quanto o modelo de autômato celular foram bem duplicados por esses alunos; eles ainda não conseguiram fazer melhor. Existem modelos complexos ajustados à sequência inteira - mas todos sabem que provavelmente são inúteis.

Dez anos depois, as estrelas começam a piscar novamente.

Dentro da recepção dos primeiros 128 bits, fica claro que a Segunda Grade pode se encaixar em pequenos movimentos no espaço 3+2 dimensional inferido, mas não se parece em nada com o estado sucessor de nenhuma das teorias dominantes de autômatos celulares. Muita alegria se segue, e os físicos começam a trabalhar para induzir que tipo de física dinâmica pode governar os objetos vistos no espaço 3+2 dimensional. Muito trabalho nesse sentido já foi feito, apenas especulando sobre que tipo de forças equilibradas poderiam dar origem aos objetos na Primeira Grade, se esses objetos fossem estáticos - mas agora parece que nem todos os objetos são estáticos. Como a maioria dos físicos supôs - teorias estaticamente equilibradas pareciam artificiais.

Muitas equações elegantes são formuladas para descrever os objetos dinâmicos no espaço 3+2 dimensional sendo projetados na Primeira e Segunda Grades. Algumas equações são mais elegantes do que outras; algumas são mais precisamente preditivas (em retrospecto, infelizmente) da Segunda Grade. Um grupo de físicos brilhantes, que se isolou cuidadosamente e olhou apenas para as primeiras 32 linhas da Segunda Grade, produz equações que parecem elegantes para eles - e as equações também se saem bem na previsão das próximas 224 linhas. Esta se torna a suposição dominante.

Mas essas equações são subespecificadas; elas não parecem ser suficientes para criar um universo. Uma pequena indústria caseira surge na tentativa de adivinhar que tipo de leis poderiam completar as que foram supostas.

Quando a Terceira Grade chega, dez anos após a Segunda Grade, ela fornece informações sobre segundas derivadas, forçando uma grande modificação da teoria “incompleta, mas boa”. Mas a teoria não se sai tão mal, considerando todas as coisas.

A Quarta Grade não adiciona muito à imagem. As terceiras derivadas não parecem importantes para a física 3+2 inferida das Grades.

A Quinta Grade parece quase exatamente como se espera que pareça.

E a Sexta Grade, e a Sétima Grade.

(Ah, e toda vez que alguém neste mundo tenta construir uma IA realmente poderosa, o hardware do computador derrete espontaneamente. Isso não é realmente importante para a história, mas preciso postular isso para ter pessoas humanas por perto, em carne e osso, por setenta anos.)

Minha moral?

Que mesmo Einstein não chegou a um milhão de anos-luz de fazer uso eficiente dos dados sensoriais.

Riemann inventou suas geometrias antes de Einstein ter um uso para elas; a física do nosso universo não é tão complicada em um sentido absoluto. Uma superinteligência Bayesiana, conectada a uma webcam, inventaria a Relatividade Geral como uma hipótese - talvez não a hipótese dominante, comparada à mecânica Newtoniana, mas ainda uma hipótese sob consideração direta - no momento em que tivesse visto o terceiro

quadro de uma maçã caindo. Poderia adivinhar a partir do primeiro quadro, se visse a estática de uma folha de grama dobrada.

Nós pensaríamos nisso. Nossa civilização, isto é, com dez anos para analisar cada quadro. Certamente, se o QI médio fosse 140 e os Einsteins fossem comuns, nós pensaríamos.

Mesmo se fôssemos inteligências de nível humano em um tipo diferente de física - mentes que nunca viram um espaço 3D projetado em uma grade 2D - ainda pensaríamos na hipótese 3D→2D. Nossos matemáticos ainda teriam inventado espaços vetoriais e projeções.

Mesmo se nunca tivéssemos visto uma bola de bilhar acelerando, nossos matemáticos teriam inventado o cálculo (por exemplo, para problemas de otimização).

Caramba, pense em toda a matemática maluca que foi inventada aqui na nossa Terra.

Ocasionalmente, encontro pessoas que dizem algo como: "Há um limite teórico sobre o quanto você pode deduzir sobre o mundo exterior, dada uma quantidade finita de dados sensoriais."

Sim. Há. O limite teórico é que cada vez que você vê 1 bit adicional, não se pode esperar que ele elimine mais da metade das hipóteses restantes (metade da massa de probabilidade restante, na verdade). E que uma mensagem redundante não pode transmitir mais informações do que a versão compactada de si mesma. Nem um bit pode transmitir qualquer informação sobre uma quantidade com a qual ele tem correlação exatamente zero entre os mundos prováveis que você imagina.

Mas nada do que eu descrevi nesta civilização humana fazendo, sequer começa a se aproximar dos limites teóricos estabelecidos pelo formalismo da indução de Solomonoff. Não se aproxima da imagem que você poderia obter se pudesse pesquisar todas as hipóteses computáveis, ponderadas por sua simplicidade, e fazer atualizações Bayesianas em todas elas.

Para ver o limite teórico da informação extraível, imagine que você tem poder computacional infinito e simula todos os universos possíveis com física simples, procurando por universos que contenham Terras embutidas neles - talvez dentro de uma simulação - onde algum processo faz as estrelas piscarem na ordem observada. Qualquer bit na mensagem - ou qualquer ordem de seleção de estrelas, aliás - que contenha a menor correlação (em todos os universos computáveis possíveis, ponderados pela simplicidade) com qualquer elemento do ambiente lhe dá informações sobre o ambiente.

A indução de Solomonoff, tomada literalmente, criaria infinitos seres sencientes, presos dentro dos cálculos. Todos os seres sencientes computáveis possíveis, na verdade. O que dificilmente parece ético. Então, vamos ficar felizes por isso ser apenas um formalismo.

Mas meu ponto é que o "limite teórico de quanta informação você pode extrair de dados sensoriais" está muito acima do que eu descrevi como o triunfo de uma civilização de físicos e criptógrafos.

Certamente não é nada parecido com um humano olhando para uma maçã caindo e pensando: "Hum, me pergunto por que isso aconteceu?"

As pessoas parecem dar um salto de "Isso é 'limitado'" para "O limite deve ser uma quantidade de aparência razoável na escala com a qual estou acostumado." A potência de uma supernova é "limitada", mas eu não aconselharia tentar se proteger de uma com um macacão Nomex retardador de chamas.

Ninguém - nem mesmo uma superinteligência Bayesiana - jamais chegará remotamente perto de fazer uso eficiente de suas informações sensoriais...

... é o que eu gostaria de dizer, mas não confio na minha capacidade de definir limites para as habilidades das superinteligências Bayesianas.

(Embora eu apostasse dinheiro nisso, se houvesse alguma maneira de julgar a aposta. Só não com probabilidades muito extremas.)

A história continua:

Milênios depois, quadro após quadro, ficou claro que alguns dos objetos na representação estão estendendo tentáculos para mover outros objetos e configurando cuidadosamente outros tentáculos para fazer sinais específicos. Eles estão tentando nos ensinar a dizer “pedra”.

Parece que os remetentes da mensagem subestimaram vastamente nossa inteligência. Do que podemos deduzir que os próprios alienígenas não são tão brilhantes. E essas crianças desajeitadas podem mudar a luminosidade de nossas estrelas? Tanto poder e tanta estupidez parecem uma combinação perigosa.

Nossos psicólogos evolucionistas começam a extrapolar possíveis cursos de evolução que poderiam produzir tais alienígenas. Um forte argumento é feito para que eles tenham evoluído assexuadamente, com trocas ocasionais de material genético e conteúdo cerebral; esta parece ser a rota mais plausível pela qual criaturas tão estúpidas ainda conseguiriam construir uma civilização tecnológica. Seus Einsteins podem ser nossos alunos de graduação, mas eles ainda poderiam coletar dados científicos suficientes para fazer o trabalho eventualmente, em dezenas de seus milênios, talvez.

A física inferida do universo 3+2 não é totalmente conhecida neste momento; mas parece certo permitir computadores muito mais poderosos do que nossos quânticos. Estamos razoavelmente certos de que nosso próprio universo está sendo executado como uma simulação em tal computador. A humanidade decide não procurar por bugs na simulação; não queremos nos desligar acidentalmente.

Nossos psicólogos evolucionistas começam a adivinhar a psicologia dos alienígenas e planejar como poderíamos persuadi-los a nos tirar da caixa. Não é difícil em um sentido absoluto - eles não são muito brilhantes - mas temos que ser muito cuidadosos...

Temos que fingir ser estúpidos também; não queremos que eles percebam seu erro.

Mas só um milhão de anos depois eles nos dizem como enviar um sinal de volta.

Neste ponto, a maior parte da espécie humana está em suspensão criônica, a temperaturas de hélio líquido, sob proteção contra radiação. Toda vez que tentamos construir uma IA ou um dispositivo nanotecnológico, ele derrete. Então a humanidade espera e dorme. A Terra é administrada por uma equipe mínima de nove supergênios. Clones, que trabalham bem juntos, sob a supervisão de certas salvaguardas de computador.

Mais cem milhões de seres humanos nascem nessa equipe, envelhecem e entram em suspensão criônica antes de terem a chance de começar a implementar lentamente os planos feitos há eras...

Da perspectiva dos alienígenas, levamos trinta de seus equivalentes a minutos para aprender, de forma inocente, sobre sua psicologia, persuadi-los com muito cuidado a nos dar acesso à Internet, seguido por cinco minutos para descobrir inocentemente seus protocolos de rede, e depois um cracking trivial cuja única dificuldade era um disfarce inocente. Lemos um punhado de artigos de física (bit por bit) de seu equivalente ao arXiv, aprendendo muito mais com seus experimentos do que eles. (A equipe da Terra gerou mais vinte Einsteins naquela geração.)

Então, decifremos o equivalente deles do problema de dobramento de proteínas em cerca de um século e fizemos alguma engenharia simulada em sua física simulada. Enviamos mensagens (codificadas esteganograficamente até que nossos servidores invadidos as decodificassem) para laboratórios que faziam seu equivalente de sequenciamento de DNA e síntese de proteínas. Encontramos um idiota desavisado, demos a ele uma história plausível e o equivalente a um milhão de dólares em dinheiro de monopólio computacional hackeado, e dissemos a ele para misturar alguns frascos que recebeu pelo correio. Equivalentes de proteínas que se auto-montaram em nanomáquinas de primeiro estágio, que construíram as nanomáquinas de segundo estágio, que construíram as nanomáquinas de terceiro estágio... e então finalmente pudemos começar a fazer as coisas em uma velocidade razoável.

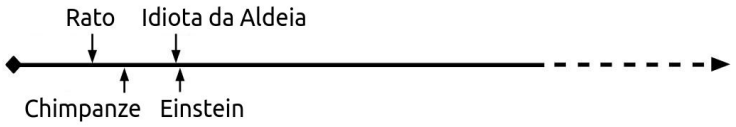
Três de seus dias, no total, desde que começaram a falar conosco. Meio bilhão de anos, para nós.

Eles nunca suspeitaram de nada. Eles não eram muito espertos, você vê, mesmo antes de considerar sua taxa de tempo mais lenta. Seus equivalentes primitivos de racionalistas andavam por aí dizendo coisas como: “Há um limite para a quantidade de informação que você pode extrair de dados sensoriais.” E eles nunca perceberam o que significava sermos mais inteligentes do que eles e pensávamos mais rápido.

254 - Meu modelo da infância

Quando dou palestras sobre a explosão da inteligência, frequentemente desenho um gráfico da "escala de inteligência" como aparece na vida cotidiana:

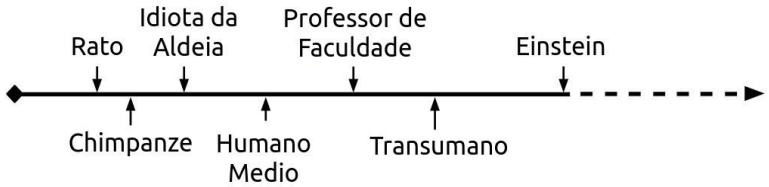
Mas esta é uma visão bastante paroquial da inteligência. Claro, na vida cotidiana, lidamos socialmente apenas com outros humanos—apenas outros humanos são parceiros no grande jogo—e, portanto, só encontramos mentes de inteligências que variam do idiota da aldeia ao Einstein. Mas o que realmente precisamos para falar sobre Inteligência Artificial ou ótimos teóricos da racionalidade é esta escala de inteligência:



A distância do "idiota da vila" ao "Einstein" é pequena no espaço dos designs de cérebro. Einstein e o idiota da aldeia têm ambos um córtex pré-frontal, um hipocampo, um cerebelo...

Talvez Einstein tenha algumas diferenças genéticas menores em relação ao idiota da aldeia, ajustes no motor. Mas a distância de design de cérebro entre Einstein e o idiota da aldeia não é nada remotamente semelhante à distância de design de cérebro entre o idiota da aldeia e um chimpanzé. Um chimpanzé não conseguiria distinguir entre Einstein e o idiota da aldeia, e nossos descendentes podem não ver muita diferença também.

Carl Shulman observou que alguns acadêmicos que falam sobre transumanismo parecem usar a seguinte escala de inteligência:



Douglas Hofstadter realmente disse algo assim, no Singularity Summit de 2006. Ele olhou para o meu diagrama mostrando o "idiota da vila" ao lado de "Einstein" e disse: "Isso parece errado para mim; acho que

Einstein deveria estar muito mais à direita."

Fiquei sem palavras. Especialmente porque era Douglas Hofstadter, um dos meus heróis de infância. Isso revelou uma lacuna cultural que eu nunca imaginei que existisse.

Veja, para mim, o que você encontraria no lado direito da escala era um Cérebro de Júpiter. Einstein não tinha literalmente um cérebro do tamanho de um planeta.

No lado direito da escala, você encontraria o Pensador Profundo—a versão original de Douglas Adams, obrigado, não o jogador de xadrez. O computador tão inteligente que, mesmo antes de seus bancos de dados estupendos serem conectados, quando foi ligado pela primeira vez, começou com "Penso, logo existo" e chegou a deduzir a existência de pudim de arroz e imposto de renda antes que alguém conseguisse desligá-lo.

No lado direito da escala, você encontraria os Anciãos de Arísia, super mentes galácticas, cérebros Matrioska e a melhor classe de Deuses. No extremo direito da escala, o Velho e a Praga.

Não o maldito Einstein.

Tenho certeza de que Einstein era muito inteligente para um humano. Tenho certeza de que um Veículo de Sistemas Gerais acharia isso muito fofo dele.

Chamo isso de "lacuna cultural" porque fui apresentado ao conceito de um Cérebro de Júpiter aos doze anos.

Tudo isso, claro, é a falácia lógica da generalização a partir de evidências fictícias.

Mas é um exemplo de por que—falácia lógica ou não—suspeito que ler ficção científica tem um efeito útil no futurismo. Às vezes, a alternativa a um conhecimento fictício de mundos fora do seu próprio é ter uma mentalidade absolutamente presa em uma era: um mundo onde humanos existem, e sempre existiram, e sempre existirão.

O universo tem 13,7 bilhões de anos, pessoal! *Homo sapiens sapiens* existe há apenas cem mil anos ou algo assim!

Por outro lado, conheci algumas pessoas que nunca leram ficção científica, mas que parecem capazes de imaginar fora do seu próprio mundo. E há fãs de ficção científica que não entendem isso. Gostaria de saber o que é "isso", para poder engarrafá-lo.

No ensaio anterior, quis falar sobre o uso eficiente das evidências, ou seja, Einstein era fofo para um humano, mas em um sentido absoluto ele era tão eficiente quanto o Departamento de Defesa dos EUA.

Então tive que falar sobre uma civilização que incluía milhares de Einsteins, pensando por décadas. Porque se eu tivesse apenas representado uma superinteligência bayesiana em uma caixa, olhando para uma webcam, as pessoas pensariam: "Mas… como ela sabe interpretar uma imagem 2D?" Elas não se colocariam no lugar da mera máquina, mesmo que fosse chamada de "superinteligência bayesiana"; não aplicariam nem mesmo sua própria criatividade ao problema do que você poderia extrair ao olhar para uma grade de bits.

Seria apenas um fantasma em uma caixa, que por acaso era chamado de "superinteligência bayesiana". O fantasma não foi informado sobre como interpretar a entrada de uma webcam; então, em seu modelo mental, o fantasma não sabe.

Quanto a saber se é realista supor que uma superinteligência bayesiana pode "fazer tudo isso"…, ou seja, as coisas que me ocorreram ao sentar para resolver o problema, escrevendo a história à medida que avançava…

Bem, deixe-me colocar desta forma: Lembre-se de como Jeffreyssai apontou que se a experiência de ter uma ideia importante não leva mais de 5 minutos, isso teoricamente lhe dá tempo para 5.760 ideias por mês? Supondo que você durma 8 horas por dia e não tenha ideias importantes enquanto dorme, é claro.

Agora, os humanos não conseguem se usar tão eficientemente. Mas os humanos não são adaptados para a tarefa de pesquisa científica. Os humanos são adaptados para perseguir cervos pela savana, atirar lanças neles, cozinhá-los e, em seguida—esta é provavelmente a parte que consome mais cérebro—argumentar habilmente que merecem receber uma parte maior da carne.

É incrível que Albert Einstein conseguiu redirecionar um cérebro assim para a tarefa de fazer física. Isso merece aplausos. Merece mais do que aplausos, merece um lugar no Livro dos Recordes do Guinness. Como construir com sucesso o carro mais rápido já feito inteiramente de gelatina.

Quão mal o deus idiota e cego (evolução) projetou realmente o cérebro humano?

Isso é algo que só pode ser compreendido através de muito estudo da ciência cognitiva, até que o horror completo comece a surgir em você.

Todos os vieses que discutimos aqui devem ser pelo menos uma dica.

Da mesma forma, o fato de que o cérebro humano deve usar todo o seu poder e concentração, com trilhões de sinapses disparando, para multiplicar dois números de três dígitos sem papel e lápis.

Einstein não fez uso eficiente de seus dados sensoriais, nem seu cérebro fez uso eficiente dos disparos de seus neurônios.

Claro, tenho certos motivos ulteriores para dizer tudo isso. Mas também é importante entender que, anos atrás, quando comecei a ser um racionalista, o ideal impossível e inatingível de inteligência que me inspirou nunca foi Einstein.

Carl Schurz disse:

Ideais são como estrelas. Você não conseguirá tocá-los com as mãos. Mas, como o marinheiro no deserto das águas, você os escolhe como seus guias e, ao segui-los, alcançará seu destino.

Então agora você teve um vislumbre de um dos meus grandes modelos de infância—meu sonho de uma IA. Apenas o sonho, é claro, a realidade não estando disponível. Alcancei esse sonho, uma vez.

E isso me ajudou em algum grau e me prejudicou em algum grau.

Pois alguns ideais são como sonhos: vêm de dentro de nós, não de fora. Mentor de Arísia procedeu da imaginação de E. E. “doc” Smith, não de algo real. Se você imagina o que uma superinteligência bayesiana diria, é apenas sua própria mente falando. Não é como uma estrela, que você pode seguir de fora.

Você tem que adivinhar onde estão seus ideais, e se adivinhar errado, você se desvia.

Mas não limite seus ideais a meras estrelas, a meros humanos que realmente existiram, especialmente se nasceram mais de cinquenta anos antes de você e estão mortos. Cada geração sucessiva tem a chance de fazer melhor. Limitar seus ideais apenas a humanos, especialmente mortos, é limitar-se ao que já foi alcançado. Você se perguntará: “Ouso fazer isso, que Einstein não pôde fazer? Isso não é *lese-majesté*?” Bem, se Einstein tivesse se sentado perguntando-se: “Estou autorizado a fazer melhor do que Newton?” ele não teria chegado onde chegou. Este é o problema de seguir estrelas; na melhor das hipóteses, isso te leva até a estrela.

Sua era te apoia mais do que você percebe, em suposições inconscientes, em uma tecnologia de mente sutilmente melhorada. Einstein era um cara legal, mas ele falava um monte de bobagens sobre um Deus impessoal, mostrando o quanto ele entendia a arte do pensamento cuidadoso em um nível mais alto de abstração do que seu próprio campo. Pode parecer menos sacrílego pensar assim se você tiver pelo menos uma supermente galáctica imaginária para comparar com Einstein, de modo que ele não seja a extremidade direita da sua escala de inteligência.

Se você tentar fazer apenas o que parece humanamente possível, pedirá muito pouco de si mesmo. Quando imagina alcançar um objetivo mais alto e inconveniente, todas as razões convenientes pelas quais

isso é “impossível” vêm prontamente à mente.

Os modelos mais importantes são os sonhos: eles vêm de dentro de nós mesmos. Sonhar com algo menor do que o que você concebe como perfeição é usar menos do que o poder total da parte de você mesmo que sonha.

255 - Os Superpoderes de Einstein

Existe uma tendência generalizada de falar (e pensar) como se Einstein, Newton e figuras históricas semelhantes tivessem superpoderes - algo mágico, algo sagrado, algo além do mundano. (Lembre-se, há muitas mais formas de venerar algo do que acender velas ao redor de seu altar.)

Eu também pensava assim inconscientemente, especialmente em relação a Einstein, até que a leitura de "O Fim do Tempo" de Julian Barbour me curou disso. [1]

Barbour expôs a história da física anti-epifenomenal e o Princípio de Mach; descreveu as controvérsias históricas que precederam Mach - tudo isso que estava por trás de Einstein e era conhecido por ele, quando Einstein abordou seu problema...

E talvez eu esteja apenas imaginando coisas - lendo demais de mim mesmo no livro de Barbour - mas achei que ouvi Barbour gritando muito silenciosamente, codificado entre as linhas educadas:

"O que Einstein fez não é mágica, pessoal! Se vocês todos apenas olhassem para como ele realmente fez isso, em vez de cair de joelhos e adorá-lo, talvez então vocês também pudessem fazê-lo!"

(Barbour não disse isso realmente. Não aparece no texto do livro. Não é uma citação de Julian Barbour e não deve ser atribuída a ele. Obrigado.)

Talvez eu esteja enganado, ou extrapolando demais... mas meio que suspeito que Barbour uma vez tentou explicar às pessoas como você avança na direção de Einstein para obter física atemporal; e eles zombaram com desdém e disseram: "Oh, você acha que é Einstein, é?"

O Índice de Maluquice de John Baez, item 18:

10 pontos para cada comparação favorável de si mesmo com Einstein, ou afirmação de que a relatividade especial ou geral são fundamentalmente equivocadas (sem boas evidências).

Item 30:

30 pontos por sugerir que Einstein, em seus últimos anos, estava tateando em direção às ideias que você agora defende.

Barbour nunca se incomoda em se comparar a Einstein, é claro; nem apela a Einstein em apoio à física atemporal. Menciono esses itens no Índice de Maluquice para mostrar quantas pessoas se comparam a Einstein, e o que a sociedade geralmente pensa delas.

O maluco vê Einstein como algo mágico, então eles se comparam a Einstein como forma de se elogiar como mágicos; eles pensam que Einstein tinha superpoderes e pensam que têm superpoderes, daí a comparação.

Mas é apenas o outro lado da mesma moeda, pensar que Einstein é sagrado, e o maluco não é sagrado, portanto, eles cometeram blasfêmia ao se compararem a Einstein.

Suponha que um jovem físico brilhante diga: "Admiro o trabalho de Einstein, mas pessoalmente, es-

pero fazer melhor." Se alguém ficar chocado e disser: "O quê! Você não realizou nada remotamente parecido com o que Einstein fez; o que o faz pensar que é mais inteligente que ele?" então eles são o outro lado da moeda do maluco.

O problema subjacente é confundir status social e potencial de pesquisa.

Einstein tem um status social extremamente alto: devido a seu histórico de realizações; devido a como ele fez isso; e porque ele é o físico cujo nome até o público lembra, que trouxe honra à própria ciência.

E tendemos a misturar fama com outras quantidades, e tendemos a atribuir o comportamento das pessoas a disposições em vez de situações.

Então há essa tendência de pensar que Einstein, mesmo antes de ser famoso, já tinha uma disposição inerente para ser Einstein - um potencial tão raro quanto sua fama e tão mágico quanto seus feitos. De modo que se você afirma ter o potencial para fazer o que Einstein fez, é o mesmo que reivindicar o posto de Einstein, elevando-se muito acima de seu status atribuído na tribo.

Não estou formulando isso bem, mas, estou tentando dissecar um pensamento confuso: Einstein pertence a um magistério separado, o magistério sagrado. O magistério sagrado é distinto do magistério mundano; você não pode se propor a ser Einstein da mesma forma que pode se propor a ser um professor titular ou um CEO. Apenas seres com potencial divino podem entrar no magistério sagrado - e então é apenas cumprir um destino que eles já têm. Então, se você diz que quer superar Einstein, você está alegando já fazer parte do magistério sagrado - você afirma ter a mesma aura de destino com que Einstein nasceu, como um direito de nascença real...

"Mas Eliezer," você diz, "certamente nem todos podem se tornar Einstein."

Você quer dizer, nem todos podem fazer melhor que Einstein.

"Hum... sim, é isso que eu quis dizer."

Bem... no mundo moderno, você pode estar correto. Você provavelmente deve se lembrar que eu sou um transumanista, andando por aí olhando para as pessoas pensando: "Sabe, é uma droga que nem todo mundo tenha o potencial de fazer melhor que Einstein, e isso parece um problema solucionável." Isso colore a atitude de alguém.

Mas no mundo moderno, sim, nem todos têm o potencial de ser Einstein.

Ainda assim... como posso colocar isso...

Há uma frase que ouvi uma vez, não me lembro onde: "Apenas mais um gênio judeu." Algum poeta ou autor ou filósofo ou outro, brilhante em uma idade jovem, fazendo algo não tremendamente importante no grande esquema das coisas, não tão influente, que acabou sendo descartado como "Apenas mais um gênio judeu."

Se Einstein tivesse escolhido um ângulo de ataque inadequado ao seu problema, ou se tivesse optado por trabalhar em um problema menos significativo, ou se tivesse desistido antes de concluir sua pesquisa, ou se tivesse tomado qualquer um dos vários caminhos errados, ou se outra pessoa tivesse resolvido o problema antes dele, o brilhante Albert Einstein poderia ter se tornado apenas mais um gênio judeu.

Gênios são raros, mas não tão raros assim. Não é tão implausível reivindicar o tipo de intelecto que pode fazer com que você seja descartado como "apenas mais um gênio judeu" ou "apenas mais uma mente brilhante que nunca fez nada interessante com sua vida." O status social associado aqui não é alto o suficiente para ser sagrado, então deve parecer uma reivindicação normalmente avaliável.

Mas o que separa pessoas assim de se tornarem Einstein, suspeito, não é nenhum defeito inato de brilhantismo. São coisas como "falta de um problema interessante" - ou, para colocar a culpa onde ela pertence, "falha em escolher um problema importante." É muito fácil falhar nisso devido ao problema do pen-

samento em cache: Diga às pessoas para escolherem um problema importante e elas escolherão o primeiro resultado em cache para "problema importante" que surgir em suas mentes, como "aquecimento global" ou "teoria das cordas."

Os problemas verdadeiramente importantes são frequentemente aqueles que você nem está considerando, porque parecem impossíveis, ou, hum, realmente difíceis, ou pior de tudo, não está claro como resolver. Se você trabalhasse neles por anos, eles poderiam não parecer tão impossíveis... mas esta é uma percepção extra e incomum; o realismo ingênuo lhe dirá que problemas solúveis parecem solúveis, e problemas que parecem impossíveis são impossíveis.

Então você tem que criar um novo e valioso ângulo de ataque. A maioria das pessoas que não são alérgicas à novidade irá longe demais na outra direção e cairá em uma espiral de morte afetiva.

E então você tem que bater a cabeça no problema por anos, sem ser distraído pelas tentações de uma vida mais fácil. "A vida é o que acontece enquanto estamos fazendo outros planos," como diz o ditado, e se você quer cumprir seus outros planos, muitas vezes tem que estar pronto para recusar a vida.

A sociedade não está configurada para apoiá-lo enquanto você trabalha, também.

O ponto é, o problema não é que você precisa de uma aura de destino e a aura de destino está faltando. Se você tivesse conhecido Albert antes de ele publicar seus artigos, você não teria percebido nenhuma aura de destino nele para corresponder ao seu futuro status elevado. Ele pareceria apenas mais um gênio judeu.

Isso não é porque o direito de nascença real está oculto, mas porque simplesmente não está lá. Não é necessário. Não há um magistério separado para pessoas que fazem coisas importantes.

Digo isso, porque quero fazer coisas importantes com minha vida, e tenho um problema genuinamente importante, e um ângulo de ataque, e tenho batido minha cabeça nisso por anos, e consegui estabelecer uma estrutura de apoio para isso; e eu frequentemente encontro pessoas que, garantidamente, dizem: "É mesmo? Vejamos sua aura de destino, cara."

O que me impressionou em Julian Barbour foi uma qualidade que eu não acho que alguém saberia como fingir sem realmente tê-la: Barbour parecia ter visto através de Einstein - ele falava sobre Einstein como se tudo o que Einstein tinha feito fosse perfeitamente compreensível e mundano.

Embora mesmo tendo percebido isso, para mim ainda foi um choque, quando Barbour disse algo como: "Agora aqui está onde Einstein falhou em aplicar seus próprios métodos, e perdeu a percepção chave -". Mas o choque foi fugaz, eu conhecia a Lei: Sem deuses, sem magia, e heróis antigos são conquistas para marcar em seu espelho retrovisor.

Este ver através é algo que se tem que alcançar, uma percepção que se tem que descobrir. Você não pode ver através de Einstein apenas dizendo "Einstein é mundano!" se seu trabalho ainda parece mágica para você. Seria como declarar "A consciência deve se reduzir a neurônios!" sem ter ideia de como fazê-lo. É verdade, mas não resolve o problema.

Não vou lhe dizer que Einstein era um sujeito comum supervalorizado pela mídia, ou que, no fundo, ele era um cara comum como todos os outros. Isso seria ir longe demais. Para percorrer esse caminho, é preciso adquirir habilidades que alguns consideram… não naturais. Tenho um prazer especial em fazer coisas que as pessoas chamam de "humanamente impossíveis", porque mostra que estou crescendo.

Todavia, a maneira como você adquire poderes mágicos não é nascendo com eles, mas vendo, com um choque repentino, que eles realmente são perfeitamente normais.

Este é um princípio geral na vida.

Referências

[1] Julian Barbour, The End of Time: The Next Revolution in Physics, 1st ed. (New York: Oxford University Press, 1999).

256 - Projeto de classe

"Fazer tão bem quanto Einstein?" Jeffreyssai disse, incrédulo. "Apenas tão bem quanto Einstein? Albert Einstein foi um grande cientista de sua época, mas aquela era a época dele, não esta! Einstein não compreendia os métodos bayesianos. Ele viveu antes da descoberta dos vieses cognitivos. Não tinha compreensão científica de seus próprios processos de pensamento. Estava muito preso ao drama de rejeitar a mecânica quântica de sua época para realmente consertá-la. E embora eu reconheça que Einstein raciocinou com clareza no caso da Relatividade Geral — exceto pela questão da constante cosmológica — ele levou dez anos para fazê-lo. Lento demais!"

"Lento demais?" repetiu Taji, incrédulo.

"Lento demais! Se Einstein estivesse nesta sala de aula agora, em vez da Terra do século primeiro negativo, eu bateria nos nós de seus dedos! Vocês não tentarão fazer tão bem quanto Einstein! Vocês aspirarão a fazer melhor que Einstein, ou nem se deem ao trabalho!"

Jeffreyssai balançou a cabeça. "Bem, já dei dicas suficientes. É hora de testar suas habilidades. Agora, sei que os outros beisutsukai não pensam muito bem dos meus projetos de classe..." Jeffreyssai fez uma pausa significativa.

Brennan suspirou internamente. Já ouvira essa fala muitas vezes antes, na Conspiração Bárdica, na Conspiração Competitiva: Os outros professores acham minhas tarefas fáceis demais, vocês deveriam ser gratos, seguido por alguma tarefa ridiculamente difícil—

"Eles dizem," continuou Jeffreyssai, "que meus projetos são difíceis demais. Insanamente difíceis. Que passam do reino da loucura para o reino de Esparta. Que o próprio Laplace pegaria fogo. Eles me acusam de tentar despedaçar as almas dos meus alunos—"

Ah, droga.

"Mas há uma razão," disse Jeffreyssai, "pela qual muitos dos meus alunos alcançaram grandes feitos. E com isso não me refiro a uma alta posição na Conspiração Bayesiana. Eu esperava muito deles, e eles passaram a esperar muito deles mesmos. Então…"

Jeffreyssai olhou para seus alunos cada vez mais perturbados. "Aqui está sua tarefa. Vocês já ouviram falar de mecânica quântica e Relatividade Geral. Este é o limite da ciência Ancestral e, portanto, o limite do conhecimento público. Vocês cinco, trabalhando sozinhos, devem produzir a teoria correta da gravidade quântica. Seu prazo é um mês."

"O quê?" disseram Brennan, Taji, Styrlyn e Yin. Hiriwa lançou-lhes um olhar intrigado.

"Se tiverem sucesso," Jeffreyssai continuou, "serão promovidos a beisutsukai do segundo dan e sexto nível. Veremos se vocês aprenderam velocidade. Seu relógio começa — agora."

E Jeffreyssai saiu da sala, batendo a porta atrás de si.

"Isso é loucura!" Taji gritou.

Hiriwa olhou para Taji, confuso. “A solução não é conhecida por nós. Como você pode saber que é tão difícil?”

“Porque já sabíamos desse problema nos dias Ancestrais! Os cientistas Ancestrais trabalharam nesse problema por muito mais de um mês.”

Hiriwa deu de ombros. “Eles ainda estavam discutindo sobre múltiplos mundos também, não estavam?”

“Chega! Não há tempo!”

Os outros quatro estudantes olharam para Styrlyn, lembrando que ele supostamente tinha uma alta posição na Conspiração Cooperativa. Houve um instante de avaliação, e então Styrlyn se tornou o líder deles.

Styrlyn respirou fundo. “Precisamos de uma lista de abordagens. Escrevam todos os ângulos que puderem pensar. Independentemente — precisamos de seus componentes individuais antes de começarmos a combinar. Em cinco minutos, pedirei a cada um de vocês sua melhor ideia primeiro. Sem pensamentos desperdiçados! Vão!”

Brennan pegou uma folha e seu traçador, colocou a ponta na superfície e então pausou. Ele não conseguia pensar em nada inteligente para dizer sobre unificar a Relatividade Geral e a mecânica quântica…

Os outros alunos já estavam escrevendo.

Brennan bateu a ponta dos dedos, uma, duas, três vezes. Relatividade Geral e mecânica quântica...

Taji colocou sua primeira folha de lado, pegou outra.

Finalmente, Brennan, por falta de algo inteligente para dizer, escreveu o óbvio.

Minutos depois, quando Styrlyn encerrou o tempo, ainda era tudo o que ele havia escrito.

“Certo,” disse Styrlyn, “sua melhor ideia. Ou a ideia que você mais quer que o resto de nós considere em nossos segundos componentes. Taji, vá!”

Taji olhou suas folhas. “Ok, acho que temos que assumir que toda avenida que a ciência Ancestral estava tentando é um beco sem saída, ou eles teriam encontrado. E se isso for possível de fazer em um mês, a resposta deve ser, de certa forma, elegante. Então, sem múltiplas dimensões. Se começarmos a fazer algo que pareça que deveríamos chamar de ‘teoria das cordas’, é melhor pararmos. Talvez começar considerando como a falha em entender a decoerência poderia ter desviado a ciência Ancestral na quantização da gravidade.”

“O oposto da loucura é loucura,” disse Hiriwa. “Vamos fingir que a ciência Ancestral nunca existiu.”

“Sem críticas ainda!” disse Styrlyn. “Hiriwa, sua sugestão?”

“Livre-se dos infinitos,” disse Hiriwa, “extirpe o que os permite. Não deveria ser uma questão de inteligência com integrais. Uma representação que permite o infinito deve ser falsa em relação aos fatos.”

“Yin.”

“Sabemos pelo senso comum,” disse Yin, “que se saíssemos do universo, veríamos o tempo disposto de uma vez, a realidade como um cristal. Mas certa vez encontrei uma dica de que a física é atemporal em um sentido mais profundo do que isso.”

Os olhos de Yin estavam distantes, lembrando. “Anos atrás, encontrei uma cidade abandonada; estava desabitada por eras, eu acho. E atrás de uma porta cujas fechaduras estavam quebradas, esculpido em uma parede: *citação .ua sai .ei mi vimcu ty bu le mekso* fim de citação.”

Brennan traduziu: Eureka! Elimine t das equações. E escrito em Lojban, a língua sagrada da ciência, o que significava que o escritor desconhecido acreditava ser verdade.

"A 'física atemporal' da qual todos ouvimos rumores," disse Yin, "pode ser atemporal em um sentido muito literal."

"Minha própria contribuição," disse Styrlyn. "A física quântica que aprendemos é sobre configurações posicionais conjuntas. Parece que deveríamos conseguir decompor isso em uma representação espacialmente local, em termos de emaranhamentos distantes invariantes. Encontrar essa representação pode nos ajudar a integrar com a Relatividade Geral, cuja curvatura é local."

"Uma perspectiva estranhamente individualista," murmurou Taji, "para alguém da Conspiração Cooperativa."

Styrlyn balançou a cabeça. "Você nos entende mal, então. A primeira lição que aprendemos é que grupos são feitos de pessoas… não, não há tempo para política. Brennan!"

Brennan deu de ombros. "Não muito, receio, apenas o óbvio. A massa-energia inercial sempre foi observada como igual à massa-energia gravitacional, e Einstein mostrou que elas eram necessariamente a mesma coisa. Então, por que a 'energia' que é um autovalor do hamiltoniano quântico é necessariamente a mesma 'energia' que aparece nas equações da Relatividade Geral? Por que o espaço-tempo deveria curvar na mesma taxa que as pequenas setas giram?"

Houve uma breve pausa.

Yin franziu a testa. "Isso parece óbvio demais. A ciência Ancestral não teria descoberto isso já?"

"Esqueça que a ciência Ancestral existiu," disse Hiriwa. "A questão permanece: precisamos da resposta, seja ela conhecida em tempos antigos ou não. Não pode ser coincidência."

Os olhos de Taji estavam abstratos. "Talvez fosse possível mostrar que uma exceção à igualdade violaria alguma lei de conservação…"

"Não é para onde Brennan apontou," interrompeu Hiriwa. "Ele não pediu uma prova de que devem ser iguais, dado algum princípio atraente; ele pediu uma visão na qual os dois são um, e não podem ser divididos nem conceitualmente, como foi realizado para a massa-energia inercial e a massa-energia gravitacional. Pois devemos assumir que a beleza do todo surge das leis fundamentais, e não o contrário. Reformulação justa?"

"Reformulação justa," respondeu Brennan.

O silêncio reinou por trinta e sete segundos, enquanto os cinco ponderavam as cinco sugestões.

"Eu tenho uma ideia…"

Interlúdio: uma explicação técnica da explicação técnica

Como Jaynes enfatiza, os teoremas da teoria da probabilidade bayesiana são apenas isso - teoremas matemáticos que seguem, inevitavelmente, os axiomas bayesianos. [1]. Pode-se pensar ingenuamente que não haveria controvérsia sobre teoremas matemáticos. Mas em quais situações os teoremas se aplicam? Como usamos os teoremas em problemas reais? A Explicação Intuitiva tenta evitar controvérsias, mas a Explicação Técnica entra deliberadamente nas pás giratórias do helicóptero. Francamente, o raciocínio na Explicação Técnica não representa o consenso unânime de toda a comunidade planetária de pesquisadores bayesianos. Pelo menos, ainda não.

Onde a Explicação Intuitiva se concentrou em fornecer uma compreensão firme dos fundamentos bayesianos, Uma Explicação Técnica da Explicação Técnica constrói, sobre uma base bayesiana, teses sobre a racionalidade humana e a filosofia da ciência. A Explicação Técnica da Explicação Técnica é assim chamada porque começa com esta pergunta:

"Qual é a diferença entre uma compreensão técnica e uma compreensão verbal?"

Quando criança, eu lia livros de física popular e me imaginava conhecedor; eu pensava saber que o som era ondas de ar, que a luz era ondas de eletromagnetismo, que a matéria era ondas de amplitudes de probabilidade complexas. Quando cresci, li as Palestras de Feynman sobre Física e reservei um tempo para entender "a equação de onda." [2] E então percebi que, até aquele ponto, eu não havia entendido ou acreditado que "o som é ondas" da maneira como um físico quer dizer e acredita nessa frase.

Então essa é a diferença entre uma compreensão técnica e uma compreensão verbal.

Você acredita nisso? Se sim, você deveria ter aplicado o conhecimento e dito: "Mas por que você não deu uma explicação técnica em vez de uma explicação verbal?"

Visualize a densidade de probabilidade ou a massa de probabilidade - a probabilidade como um pedaço de argila que você deve distribuir sobre os resultados possíveis.

Digamos que haja uma pequena luz que pode piscar em vermelho, azul ou verde cada vez que você pressiona um botão. A luz pisca em uma e apenas uma cor a cada pressionamento do botão; as possibilidades são mutuamente exclusivas. Você está tentando prever a cor do próximo flash. Em cada tentativa, você tem um peso de argila, a massa de probabilidade, que você tem que distribuir sobre as possibilidades vermelho, verde e azul. Você pode colocar um quarto de sua argila na possibilidade verde, um quarto de sua argila na possibilidade azul e metade de sua argila na possibilidade vermelha - como atribuir probabilidades de 25% ao verde, 25% ao azul e 50% ao vermelho. A metáfora é que a probabilidade é um recurso conservado, a ser distribuído com moderação. Se você acha que o azul tem mais probabilidade de piscar no próximo experimento, você pode atribuir uma probabilidade maior ao azul, mas você tem que tirar a massa de probabilidade das outras hipóteses - talvez roubar um pouco de argila do vermelho e adicioná-la ao azul. Você nunca pode obter mais argila. Suas probabilidades não podem somar mais do que 1,0 (100%). Você não pode prever 75% de chance de ver vermelho e 80% de chance de ver azul.

Por que você gostaria de ter cuidado com sua massa de probabilidade ou distribuí-la com moderação? Por que não espalhar a probabilidade por todo o lugar? Mudemos a metáfora da argila para o dinheiro. Você pode apostar até um dólar de dinheiro de jogo a cada pressionamento do botão. Um experimentador

está por perto e paga a você uma quantia de dinheiro real que depende de quanto dinheiro de jogo você apostou na luz vencedora. Não nos importamos como você distribuiu seu dinheiro de jogo restante sobre as luzes perdedoras. A única coisa que importa é quanto você apostou na luz que realmente venceu.

Mas devemos construir cuidadosamente a regra de pontuação usada para pagar os vencedores, se quisermos que os jogadores sejam cuidadosos com suas apostas. Suponha que o experimentador pague a cada jogador dinheiro real igual ao dinheiro de jogo apostado na cor vencedora. Sob essa regra de pontuação, se você observar que o vermelho aparece seis vezes em dez, sua melhor estratégia é apostar, não 60 centavos no vermelho, mas o dólar inteiro no vermelho, e você não se importa com as frequências do azul e do verde. Por quê? Digamos que o azul e o verde apareçam cerca de duas vezes em dez. E suponha que você aposte 60 centavos no vermelho, 20 centavos no azul e 20 centavos no verde. Nesse caso, seis vezes em dez você ganharia 60 centavos e quatro vezes em dez você ganharia 20 centavos, para um pagamento médio de 44 centavos. Sob essa regra de pontuação, faz mais sentido alocar o dólar inteiro para o vermelho e ganhar um dólar inteiro seis vezes em dez. Quatro vezes em dez você não ganharia nada. Seu pagamento médio seria de 60 centavos.

Se escrevêssemos a função para o pagamento, seria Pagamento = P(vencedor), onde P(vencedor) é a quantidade de dinheiro de jogo que você apostou na cor vencedora naquela rodada. Se escrevêssemos a função para o pagamento esperado, dada regra de pagamento, essa função seria:

$$Expectativa(Prêmio) = \sum_{cores} P(cor) \times F(cor)$$

P(cor) é a quantidade de dinheiro que você apostou em uma cor e F(cor) é a frequência com que essa cor vence.

Suponha que as frequências reais das luzes sejam 30% azul, 20% verde e 50% vermelho. E suponha que em cada rodada eu aposte 40% no azul, 50% no verde e 10% no vermelho. Eu receberia 40 centavos 30% das vezes, 50 centavos 20% das vezes e 10 centavos 50% das vezes, para um pagamento médio de $ 0,12 + $ 0,10 + $ 0,05 ou $ 0,27. Isso é:

P(cor) = dinheiro de jogo atribuído a essa cor

F(cor) = frequência com que essa cor vence

Pagamento = P(vencedor) = quantidade de dinheiro de jogo alocado para a cor vencedora.

Frequências reais de vitória:

F(azul) = 30%

F(verde) = 20%

F(vermelho) = 50%:

No longo prazo, o vermelho vence 50% das vezes, o verde vence 20% das vezes e o azul vence 30% das vezes. Portanto, nosso pagamento médio em cada rodada é 50% do pagamento se o vermelho vencer, mais 20% do pagamento se o verde vencer, mais 30% do pagamento se o azul vencer.

O pagamento é uma função da cor vencedora e do esquema de apostas. Queremos calcular o pagamento médio, dado um esquema de apostas e as frequências em que cada cor vence. O termo matemático para esse tipo de cálculo, tomando uma função de cada caso e ponderando-a pela frequência desse caso, é uma expectativa. Assim, para calcular nosso pagamento esperado, calcularíamos:

$$Expectativa(Prêmio) = \sum_{cores} P(cor) \times F(cor)$$

$$= P(azul) \times F(azul)$$

$$+ P(verde) \times F(verde)$$

$$+ P(vermelho) \times F(vermelho)$$

$$= \$0,40 \times 30\% + \$0,50 \times 20\% + \$0,10 \times 50\%$$

$$= \$0,12 + \$0,10 + \$0,05$$

$$= \$0,27$$

Com este esquema de apostas, ganharei, em média, cerca de 27 centavos por rodada.

Aloquei meu dinheiro de jogo de uma forma grosseiramente arbitrária, e surge a pergunta: posso aumentar meu pagamento esperado alocando meu dinheiro de jogo com mais sabedoria?

Dada a regra de pontuação fornecida, eu maximizo meu pagamento esperado alocando meu dólar inteiro para o vermelho. Apesar do meu pagamento esperado de 50 centavos por rodada, a luz pode realmente piscar verde, azul, azul, verde, verde e eu receberia um pagamento real de zero. No entanto, a chance da luz aparecer não vermelha em cinco rodadas sucessivas é de aproximadamente 3%. Compare o jogo de cartas vermelho / azul em Incerteza Legal.

Uma regra de pontuação adequada é uma regra para pontuar apostas para que você maximize seu pagamento esperado apostando dinheiro de jogo que seja exatamente igual à chance de aquela cor piscar. Queremos uma regra de pontuação para que, se as luzes realmente piscarem nas frequências 30% azul, 20% verde e 50% vermelho, você possa maximizar seu pagamento médio apenas apostando 30 centavos no azul, 20 centavos no verde e 50 centavos no vermelho. Uma regra de pontuação adequada é aquela que força sua aposta ideal a relatar exatamente sua estimativa das probabilidades. (Isso também é conhecido como uma regra de pontuação estritamente adequada.) Como vimos, nem todas as regras de pontuação têm essa propriedade; e se você inventar uma regra de pontuação plausível ao acaso, ela provavelmente não terá a propriedade.

Uma regra com esta propriedade adequada é pagar um dólar menos o erro ao quadrado da aposta, em vez da própria aposta - se você apostar 30 centavos na luz vencedora, seu erro seria de 70 centavos, seu erro ao quadrado seria de 49 centavos ($0,7^2 = 0,49$), e um dólar menos seu erro ao quadrado seria de 51 centavos. [3] (Presumivelmente, seu dinheiro de jogo é denominado na raiz quadrada de centavos, de modo que o erro ao quadrado é uma soma monetária.)

Não usaremos a regra do erro ao quadrado. Os estatísticos comuns consideram o erro ao quadrado de tudo à vista, mas não os estatísticos bayesianos.

Adicionamos um novo requisito: exigimos, não apenas, uma regra de pontuação adequada, mas que nossa regra de pontuação adequada nos dê a mesma resposta, quer a apliquemos a rodadas individualmente ou combinadas. Isso é o que os bayesianos fazem em vez de tomar o erro ao quadrado das coisas; exigimos invariâncias.

Suponha que eu pressione o botão duas vezes seguidas. Existem nove resultados possíveis: verde-verde, verde-azul, verde-vermelho, azul-verde, azul-azul, azul-vermelho, vermelho-verde, vermelho-azul e vermelho-vermelho. Suponha que o verde ganhe e, em seguida, o azul ganhe. O experimentador atribuiria a primeira pontuação com base em nossas atribuições de probabilidade para P(verde1) e a segunda pontuação com base em P(azul2 | verde1). [4] Faríamos duas previsões e obteríamos duas pontuações. Nossa primeira previsão foi a probabilidade que atribuímos à cor que venceu na primeira rodada, verde. Nossa segunda previsão foi nossa probabilidade de que o azul venceria na segunda rodada, dado que o verde venceu na primeira rodada. Por que precisamos escrever P(azul2 | verde1) em vez de apenas P(azul2)? Porque você pode ter uma hipótese sobre a luz piscando que diz "azul nunca segue verde" ou "azul sempre segue verde" ou "azul segue verde com 70% de probabilidade". Se for esse o caso, após ver o verde na primeira rodada, você pode querer revisar sua previsão - mudar suas apostas - para a segunda rodada. Você sempre pode revisar suas previsões até o momento em que o experimentador pressiona o botão, usando cada pedaço de informação; mas depois que a luz pisca, é tarde demais para mudar sua aposta.

Suponha que o resultado real seja verde1 seguido por azul2. Exigimos esta invariância: devo obter a mesma pontuação total, independentemente de:

- Sou pontuado duas vezes, primeiro na minha previsão para P(verde1) e segundo na minha previsão para P(azul2 | verde1).
- Sou pontuado uma vez pela minha previsão conjunta P(verde1 e azul2).

Suponha que eu atribua uma probabilidade de 60% ao verde1 e, em seguida, a luz verde pisque. Agora devo produzir probabilidades para as cores na segunda rodada. Avalio a possibilidade azul2 e aloco a ela 25% da minha massa de probabilidade. Eis que, na segunda rodada, a luz pisca em azul. Então, na primeira rodada, minha aposta na cor vencedora foi de 60% e, na segunda rodada, minha aposta na cor vencedora foi de 25%. Mas eu também poderia, no início do experimento e após atribuir P(verde1), imaginar que a luz primeiro pisca em verde, imaginar atualizar minhas teorias com base nessa informação e então dizer qual confiança darei ao azul na próxima rodada se a primeira rodada for verde. Ou seja, eu gero as probabilidades P(verde1) e P(azul2 | verde1). Multiplicando essas duas probabilidades juntas, obteríamos a probabilidade conjunta, P(verde1 e azul2) = 15%.

Um experimento duplo tem nove resultados possíveis. Se eu gerar nove probabilidades para P(verde1,verde2), P(verde1,azul2),..., P(vermelho1,azul2), P(vermelho1,vermelho2), a massa de probabilidade deve somar no máximo um. Estou dando previsões para nove possibilidades mutuamente exclusivas de um "experimento duplo".

Exigimos uma regra de pontuação (e talvez não se pareça com nada que um agenciador de apostas real jamais usaria) de forma que minha pontuação não mude, independentemente de considerarmos o resultado duplo como duas previsões ou uma previsão. Posso tratar a sequência de dois resultados como um único experimento, "pressione o botão duas vezes" e ser pontuado na minha previsão para P(azul2,verde1) = 15%. Ou posso ser pontuado uma vez pela minha primeira previsão P(verde1) = 60%, depois novamente na minha previsão P(azul2 | verde1) = 25%. Exigimos a mesma pontuação total em ambos os casos, para não importar como dividimos os experimentos e as previsões - a pontuação total é sempre a mesma. Esta é a nossa invariância.

Acabamos de exigir:

Pontuação[P(verde1,azul2)] = Pontuação[P(verde1)] + Pontuação[P(azul2 | verde1)].

E nós já sabemos:

P(verde1,azul2) = P(verde1) × P(azul2 | verde1).

A única regra de pontuação possível é:

Pontuação(P) = log(P).

A nova regra de pontuação é que sua pontuação é o logaritmo da probabilidade que você atribuiu ao vencedor.

A base do logaritmo é arbitrária - quer usemos o logaritmo de base dez ou o logaritmo de base dois, a regra de pontuação tem a invariância desejada. Mas devemos escolher alguma base real. Um matemático escolheria a base e; um engenheiro escolheria a base dez; um cientista da computação escolheria a base dois. Se usarmos a base dez, podemos converter para decibéis, como na Explicação Intuitiva; mas às vezes os bits são mais fáceis de manipular.

A regra de pontuação do logaritmo é adequada - ela tem seu máximo esperado quando dizemos nossas antecipações exatas; recompensa a honestidade. Se pensarmos que a luz azul tem 60% de probabilidade de piscar e calcularmos nosso pagamento esperado para diferentes esquemas de apostas, descobrimos que maximizamos nosso pagamento esperado dizendo ao experimentador "60%". (Leitores com cálculo podem verificar isso.) A regra de pontuação também dá um total invariante, independentemente de pressionar o botão duas vezes contar como "um experimento" ou "dois experimentos". No entanto, os pagamentos agora são todos negativos, pois estamos pegando o logaritmo da probabilidade e a probabilidade está entre zero e um. O logaritmo de base dez de 0,1 é -1; o logaritmo de base dez de 0,01 é -2. Tudo bem. Aceitamos que a regra de pontuação pode não se parecer com nada que um agenciador de apostas real jamais usaria. Se quiser, você pode imaginar que o experimentador tenha uma pilha de dinheiro e, no final do experimento, ele lhe concede uma quantia menos sua grande pontuação negativa. (Er, o valor mais sua pontuação negativa.) Talvez o experimentador tenha cem dólares e, no final de cem rodadas, você acumulou uma pontuação de -48, então você recebe $52 dólares.

Uma pontuação de -48 em que base? Podemos eliminar a ambiguidade na pontuação especificando unidades. Dez decibéis equivalem a um fator de 10; dez decibéis negativos equivalem a um fator de 1/10. Atribuir uma probabilidade de 0,01 ao resultado real pontuaria -20 decibéis. Uma probabilidade de 0,03 pontuaria -15 decibéis. Às vezes, podemos usar bits: 1 bit é um fator de 2, -1 bit é um fator de 1/2. Uma probabilidade de 0,25 pontuaria -2 bits; uma probabilidade de 0,03 pontuaria em torno de -5 bits.

Se você chegar a uma avaliação de probabilidade P para cada cor, com P(vermelho), P(azul), P(verde), então sua pontuação esperada é:

Pontuação(P) = log(P)

$$Expectativa(Pontuação) = \sum_{cores} P(cor) \times log(P(cor))$$

Suponha que você tivesse probabilidades de 25% vermelho, 50% azul e 25% verde. Pensemos na base 2 por um momento, para tornar as coisas mais simples. Sua pontuação esperada é:

Pontuação(vermelho) = -2 bits, pisca 25% do tempo,

Pontuação(azul) = -1 bit, pisca 50% do tempo,

Pontuação(verde) = -2 bits, pisca 25% do tempo,

Expectativa(Pontuação) = −1,5 bits.

Compare nossa regra de pontuação bayesiana com a forma comum ou coloquial de falar sobre graus de crença, onde alguém pode dizer casualmente: "Tenho 98% de certeza de que o óleo de canola contém mais gorduras ômega-3 do que o azeite de oliva". O que eles realmente querem dizer com isso é que eles se sentem 98% certos - há algo como uma pequena barra de progresso que mede a força da emoção da certeza, e essa barra de progresso está 98% cheia. E a barra de progresso emocional provavelmente não estaria exatamente 98% cheia, se tivéssemos alguma forma de medir. A palavra "98%" é apenas uma forma coloquial de dizer: "Estou quase, mas não totalmente certo". Não significa que você poderia obter o maior pagamento esperado apostando exatamente 98 centavos de dinheiro de jogo nesse resultado. Você só deve atribuir uma confiança calibrada de 98% se estiver confiante o suficiente para pensar que poderia responder a cem

perguntas semelhantes, de igual dificuldade, uma após a outra, cada uma independente das outras, e estar errado, em média, cerca de duas vezes. Manteremos o controle da frequência com que você está certo, ao longo do tempo, e se acontecer de quando você diz “90% de certeza” você está certo cerca de sete vezes em dez, então diremos que você está mal calibrado.

Se você disser “98% provável” mil vezes e for surpreendido apenas cinco vezes, ainda o criticaremos por uma calibração ruim. Você está alocando muita massa de probabilidade para a possibilidade de estar errado. Você deve dizer “99,5% provável” para maximizar sua pontuação. A regra de pontuação recompensa a calibração precisa, não encorajando nem a humildade, nem a arrogância.

Neste ponto, pode ocorrer a alguns leitores que existe uma maneira óbvia de alcançar a calibração perfeita - basta jogar uma moeda para cada pergunta sim ou não, e atribuir à sua resposta uma confiança de 50%. Você diz 50% e está certo na metade das vezes. Isso não é uma calibração perfeita? Sim. Mas a calibração é apenas um componente de nossa pontuação bayesiana; o outro componente é a discriminação.

Suponha que eu lhe faça dez perguntas sim ou não. Você não sabe absolutamente nada sobre o assunto, então em cada pergunta você divide sua massa de probabilidade cinquenta e cinquenta entre “Sim” e “Não”. Parabéns, você está perfeitamente calibrado - as respostas para as quais você disse “50% de probabilidade” eram verdadeiras exatamente na metade das vezes. Isso é verdade independentemente da sequência de respostas corretas ou de quantas respostas foram Sim. Em dez experimentos, você disse “50%” em vinte ocasiões - você disse “50%” para Sim1, Não1, Sim2, Não2, Sim3, Não3,... Em dez dessas ocasiões, a resposta estava correta, as ocasiões: Sim1, Não2, Não3,... E em dez dessas ocasiões a resposta estava incorreta: Não1, Sim2, Sim3,…

Agora dou minhas próprias respostas, me esforçando mais, tentando discriminar se “Sim” ou “Não” é a resposta correta. Atribuo 90% de confiança a cada uma das minhas respostas favoritas, e minha resposta favorita está errada duas vezes. Estou mais mal calibrado do que você. Eu disse “90%” em dez ocasiões e errei duas vezes. Da próxima vez que alguém me ouvir, pode traduzir mentalmente “90%” em 80%, sabendo que quando tenho 90% de certeza, estou certo cerca de 80% das vezes. Mas a probabilidade que você atribuiu ao resultado é 1/2 elevado à décima potência, que é 0,001 ou 1/1024. A probabilidade que atribuí ao resultado é 90% elevado à oitava potência vezes 10% elevado à segunda potência, $0{,}9^8 \times 0{,}1^2$, dando 0,004 ou 0,4%. Sua calibração é perfeita e a minha não, mas minha melhor discriminação entre respostas certas e erradas mais do que compensa. Minha pontuação final é mais alta - atribuí uma maior probabilidade conjunta ao resultado de todo o experimento. Se eu tivesse sido menos confiante e melhor calibrado, a probabilidade que atribuí ao resultado teria sido $0{,}8^8 \times 0{,}2^2$, dando 0,006 ou 6%.

É possível fazer ainda melhor? Claro. Você poderia ter adivinhado cada resposta corretamente e atribuído uma probabilidade de 99% a cada uma de suas respostas. Então, a probabilidade que você atribuiu a todo o resultado experimental seria $0{,}99^{10} \approx 90\%$.

Sua pontuação seria log(90%), que é −0,45 decibéis ou −0,15 bits. Precisamos aplicar o logaritmo para que, se eu tentar maximizar minha pontuação esperada, $\sum P \times \log(P)$, não tenho motivos para trapacear. Sem a regra do logaritmo, eu maximizaria minha pontuação esperada atribuindo toda a minha massa de probabilidade ao resultado mais provável. Além disso, sem a regra do logaritmo, minha pontuação total seria diferente dependendo se contássemos várias rodadas como vários experimentos ou como um experimento.

Uma transformação simples pode corrigir a calibração ruim diminuindo a discriminação. Se você tem o hábito de dizer “milhão para um” em 90 respostas corretas e 10 incorretas para cada cem perguntas, podemos aperfeiçoar sua calibração substituindo “milhão para um” por “nove para um”. Em contraste, não há uma maneira fácil de aumentar a discriminação (bem-sucedida). Se você costuma dizer “nove para um” em 90 respostas corretas para cada cem perguntas, posso facilmente aumentar sua discriminação alegada substituindo “nove para um” por “milhão para um”. Mas nenhuma transformação simples pode aumentar sua discriminação real de forma que sua resposta distinga 95 respostas corretas e 5 respostas incorretas. De Yates et al.: [5]

> Considerando que uma boa calibração geralmente pode ser alcançada por meio de transformações matemáticas simples (por exemplo, adicionando uma constante a cada julgamento de probabilidade), uma boa discrimi-

nação exige acesso a evidências sólidas e preditivas e habilidade para explorar essas evidências, que são difíceis de encontrar em qualquer situação prática da realidade[52].

Se você não consegue distinguir a verdade da falsidade, você pode alcançar a calibração perfeita confessando sua ignorância; mas confessar a ignorância não irá, por si só, distinguir a verdade da falsidade.

Assim, dispomos de outro falso estereótipo de racionalidade, de que a racionalidade consiste em ser humilde e modesto e confessar impotência diante do desconhecido. Essa é apenas a saída do trapaceiro, atribuindo uma probabilidade de 50% a todas as perguntas, sim ou não. Nossa regra de pontuação o incentiva a fazer melhor se puder. Se você é ignorante, confesse sua ignorância; se você está confiante, confesse sua confiança. Nós o penalizamos por ser confiante e errado, mas também o recompensamos por ser confiante e certo. Essa é a virtude de uma regra de pontuação adequada.

Suponha que eu jogue uma moeda vinte vezes. Se eu acredito que a moeda é justa, a melhor previsão que posso fazer é prever uma chance igual de cara ou coroa em cada lançamento. Se eu acredito que a moeda é justa, atribuo a mesma probabilidade a cada sequência possível de vinte lançamentos de moeda. Existem cerca de um milhão (1.048.576) sequências possíveis de vinte lançamentos de moeda, e eu tenho apenas 1,0 de massa de probabilidade para brincar. Então, eu atribuo a cada sequência individual possível uma probabilidade de $(1/2)^{20}$ - chances de cerca de um milhão para um; −20 bits ou −60 decibéis.

Fiz uma previsão experimental e obtive uma pontuação de -60 decibéis! Isso não falsifica a hipótese? Intuitivamente, não. Não jogamos uma moeda vinte vezes e vemos um resultado de aparência aleatória, então recuamos e dizemos, por que, as chances disso são de um milhão para um. Mas as chances são de um milhão para um contra ver essa sequência exata, como eu descobriria se ingenuamente previsse o mesmo resultado exato para a próxima sequência de vinte lançamentos de moeda. Não há problema em ter teorias que atribuem probabilidades minúsculas aos resultados, contanto que nenhuma outra teoria se saia melhor. Mas se alguém usasse uma hipótese alternativa para escrever a sequência exata em um envelope lacrado com antecedência, e ela atribuísse uma probabilidade de 99%, eu suspeitaria da justiça da moeda. Contanto que ela tenha selado apenas um envelope, e não um milhão.

Isso nos diz o que devemos responder com bom senso, mas não diz como a resposta do bom senso surge da matemática. Para dizer por que o bom senso está correto, precisamos integrar tudo o que foi dito até agora na estrutura da revisão bayesiana da crença. Quando terminarmos, teremos uma compreensão técnica da diferença entre uma compreensão verbal e uma compreensão técnica.

Imagine um experimento que produz um resultado inteiro entre zero e 99. Por exemplo, o experimento pode ser um contador de partículas que nos diz quantas partículas passaram em um minuto. Ou o experimento pode ser visitar o supermercado na quarta-feira, verificar o preço de um saco de 380 g. de nozes trituradas e anotar os dois últimos dígitos do preço.

Estamos testando várias hipóteses diferentes que tentam prever o resultado experimental. Cada hipótese produz uma distribuição de probabilidade sobre todos os resultados possíveis; neste caso, os inteiros entre zero e 99. As possibilidades são mutuamente exclusivas, então a massa de probabilidade na distribuição deve somar um (ou menos); não podemos prever uma probabilidade de 90% de ver 42 e também uma probabilidade de 90% de ver 43.

Suponha que haja uma hipótese precisa que prevê 90% de chance de obtermos o resultado 51. (Ou seja, a hipótese é que o supermercado geralmente precifica nozes com um preço de "X dólares e 51 centavos"). A teoria precisa apostou 90% de sua massa de probabilidade no resultado 51. Isso deixa 10% da massa de probabilidade restante para se espalhar por 99 outros resultados possíveis - todos os números entre zero e 99, exceto 51. A teoria não faz mais especificações, então espalhamos os 10% restantes da massa de probabilidade uniformemente sobre 99 possibilidades, atribuindo uma probabilidade de 1/990 para cada resultado não 51. Para facilitar a escrita, aproximaremos 1/990 como 0,1%.

52 NT. Texto original em inglês. *Whereas good calibration often can be achieved by simple mathematical transformations (e.g., adding a constant to every probability judgment), good discrimination demands access to solid, predictive evidence and skill at exploiting that evidence, which are difficult to find in any real-life, practical situation.*

Esta distribuição de probabilidades é análoga à verossimilhança ou probabilidade condicional do resultado dada a hipótese. Vamos chamá-la de distribuição de verossimilhança para a hipótese, nossa chance de ver cada resultado especificado se a hipótese for verdadeira. A distribuição de verossimilhança para uma hipótese H é uma função composta de todas as probabilidades condicionais para P(0|H) = 0,001, P(1|H) = 0,001,..., P(51|H) = 0,9,..., P(99jH) = 0,001.

A teoria precisa prevê uma probabilidade de 90% de ver 51. Que haja também uma teoria vaga, que prevê "uma probabilidade de 90% de ver um número na casa dos cinquenta".

Vendo o resultado 51, não dizemos que o resultado confirma ambas as teorias igualmente. Ambas as teorias fizeram previsões e ambas atribuíram probabilidades de 90%, e o resultado 51 confirma ambas as previsões. Mas a teoria precisa tem uma vantagem porque concentra sua massa de probabilidade em um ponto mais nítido. Se a teoria vaga não fizer mais especificações, contamos "uma probabilidade de 90% de ver um número na casa dos cinquenta" como uma probabilidade de 9% de ver cada número entre 50 e 59.

Suponha que começamos com chances iguais em favor da teoria precisa e da teoria vaga - chances de 1:1 ou 50% de probabilidade de qualquer uma das hipóteses ser verdadeira. Após ver o resultado 51, quais são as chances posteriores da teoria precisa ser verdadeira? As previsões das duas teorias são análogas às suas atribuições de verossimilhança - a probabilidade condicional de ver o resultado, dado que a teoria é verdadeira. Qual é a razão de verossimilhança entre as duas teorias? A primeira teoria alocou 90% da massa de probabilidade para o resultado exato. A teoria vaga alocou 9% da massa de probabilidade para o resultado exato. A razão de verossimilhança é 10:1. Portanto, se começássemos com chances iguais de 1:1, as chances posteriores seriam de 10:1 em favor da teoria precisa. A pressão diferencial das duas probabilidades condicionais empurrou nossa confiança anterior de 50% para uma confiança posterior de cerca de 91% de que a teoria precisa está correta. Assumindo que essas são as únicas hipóteses sendo testadas, que esta é a única evidência em consideração e assim por diante.

Por que a teoria vaga perdeu quando ambas as teorias se encaixam nas evidências? A teoria vaga é tímida; faz uma previsão ampla, protege suas apostas, permite muitas possibilidades que falsificariam a teoria precisa. Esta não é a virtude de uma teoria científica. Os filósofos da ciência nos dizem que as teorias devem ser ousadas e se submeter de bom grado à falsificação se sua previsão falhar. [6] Agora vemos por quê. A teoria precisa concentra sua massa de probabilidade em um ponto mais nítido e, portanto, se torna vulnerável à falsificação se o resultado real atingir outro lugar; mas se o resultado previsto estiver correto, a precisão tem uma tremenda vantagem de verossimilhança sobre a vagueza.

As leis da teoria da probabilidade não fornecem nenhuma maneira de trapacear, para fazer uma hipótese vaga de forma que qualquer resultado entre 50 e 59 conte para tanta confirmação favorável quanto a teoria precisa recebe, pois isso exigiria uma massa de probabilidade somando 900%. Não há como trapacear, desde que você registre sua previsão com antecedência, para não poder alegar depois que sua teoria atribui uma probabilidade de 90% a qualquer resultado que chegue. Os humanos gostam muito de fazer suas previsões depois, então o processo social da ciência requer uma previsão antecipada antes de dizermos que um resultado confirma uma teoria. Mas como os humanos podem se mover em harmonia com o caminho de Bayes, e assim exercer o poder, é uma questão separada de se a matemática funciona. Quando estamos fazendo matemática, apenas consideramos que as funções de densidade de verossimilhança são propriedades fixas de uma hipótese e a massa de probabilidade soma 1 e você nunca sonharia em fazer de outra forma.

Você pode querer reservar um momento para visualizar que, se definirmos a probabilidade em termos de calibração, o Teorema de Bayes relaciona as calibrações. Suponha que acho que a Teoria 1 tem 50% de probabilidade de ser verdadeira e acho que a Teoria 2 tem 50% de probabilidade de ser verdadeira. Suponha que eu esteja bem calibrado; quando pronuncio as palavras "cinquenta por cento", o evento acontece cerca de metade das vezes. E então eu vejo um resultado R que aconteceria cerca de nove décimos das vezes dada a Teoria 1, e cerca de nove centésimos das vezes dada a Teoria 2, e eu sei que isso é assim, e aplico o raciocínio bayesiano. Se eu fosse perfeitamente calibrado inicialmente (apesar da má discriminação de dizer 50/50), ainda estaria perfeitamente calibrado (e melhor discriminado) após dizer que minha confiança na Teoria 1 agora é de 91%. Se eu repetisse esse tipo de situação muitas vezes, estaria certo cerca de dez onze avos das vezes quando dissesse "91%". Se eu raciocinar usando as regras bayesianas e começar com antece-

dentes bem calibrados, minhas conclusões também serão bem calibradas. Isso só é verdade se definirmos a probabilidade em termos de calibração! Se “90% de certeza” for interpretado como, digamos, a força da emoção da certeza, não há razão para esperar que a emoção posterior tenha uma relação bayesiana exata com a emoção anterior.

Deixe as probabilidades anteriores serem de dez para um em favor da teoria vaga. Por quê? Suponha que nossa maneira de descrever hipóteses nos permita especificar um número preciso ou apenas especificar um primeiro dígito; podemos dizer “51”, “63”, “72” ou “na casa dos cinquenta / sessenta / setenta”. Suponha que pensemos que a resposta real é igualmente provável de ser uma resposta do primeiro tipo ou do segundo. No entanto, dado o problema, existem cem hipóteses possíveis do primeiro tipo e apenas dez hipóteses do segundo tipo. Portanto, se pensarmos que qualquer classe de hipóteses tem uma chance anterior quase igual de estar correta, temos que espalhar a massa de probabilidade anterior sobre dez vezes mais teorias precisas do que teorias vagas. A teoria precisa que prevê exatamente 51 teria, portanto, um décimo da massa de probabilidade anterior da teoria vaga que prevê um número na casa dos cinquenta. Após ver 51, as chances iriam de 1:10 em favor da teoria vaga para 1: 1, chances iguais para a teoria precisa e a teoria vaga.

Se você olhar para isso com atenção, é exatamente o que o bom senso esperaria. Você começa incerto se um fenômeno é o tipo de fenômeno que produz o mesmo resultado todas as vezes, ou se é o tipo de fenômeno que produz um resultado na casa dos X todas as vezes. (Talvez o fenômeno seja uma faixa de preço no supermercado, se você precisar de algum motivo para supor que 50–59 é uma faixa aceitável, mas 49–58 não é.) Você faz uma única medição e a resposta é 51. Bem, isso pode ser porque o fenômeno é exatamente 51, ou porque está na casa dos cinquenta. Portanto, a teoria precisa restante tem as mesmas chances que a teoria vaga restante, o que requer que a teoria vaga deve ter começado dez vezes mais provável do que essa teoria precisa, uma vez que a teoria precisa tem um ajuste mais preciso às evidências.

Se virmos apenas um número, como 51, isso não muda a probabilidade anterior de que o próprio fenômeno fosse “preciso” ou “vago”. Mas, na verdade, concentra toda a massa de probabilidade dessas duas classes de hipóteses em uma única hipótese sobrevivente de cada classe.

Claro, é um erro grave dizer que um fenômeno é preciso ou vago, um caso do que Jaynes chama de Falácia da Projeção da Mente. [7] Precisão ou vagueza é uma propriedade dos mapas, não dos territórios. Em vez disso, devemos perguntar se o preço no supermercado permanece constante ou muda. Uma hipótese do tipo “vaga” é uma boa descrição de um preço que muda. Um mapa preciso se adequará a um território constante.

Outro exemplo: você joga uma moeda dez vezes e vê a sequência cara-cara-coroa-coroa-cara:coroa-coroa-coroa-coroa-cara. Talvez você tenha começado pensando haver 1% de chance de essa moeda ser fixa. A hipótese “Esta moeda é fixa para produzir cara-cara-coroa-coroa-cara:coroa-coroa-coroa-coroa-cara” não atribui mil vezes a massa de verossimilhança ao resultado observado, em comparação com a hipótese da moeda justa? Sim. As chances posteriores de que a moeda seja fixa não vão para 10: 1? Não. A probabilidade anterior de 1% de que “a moeda seja fixa” deve cobrir todos os tipos possíveis de moeda fixa - uma moeda fixa para produzir cara-cara-coroa-coroa-cara : coroa-coroa-coroa-coroa-cara, uma moeda fixa para produzir coroa-coroa-cara-cara-coroa:cara-cara-cara-cara-coroa, etc. A probabilidade anterior de que a moeda seja fixa para produzir cara-cara-coroa-coroa-cara : coroa-coroa-coroa-coroa não é 1%, mas um milésimo de um por cento. Depois, a probabilidade posterior de que a moeda seja fixa para produzir cara-cara-coroa-coroa-cara:coroa-coroa-coroa-coroa-cara é de um por cento. Ou seja: você pensou que a moeda era provavelmente justa, mas tinha um por cento de chance de ser fixada em alguma sequência aleatória; você jogou a moeda; a moeda produziu uma sequência de aparência aleatória; e isso não lhe diz nada sobre se a moeda é justa ou fixa. Diz a você, se a moeda for fixa, a qual sequência ela está fixa.

Esta parábola ajuda a ilustrar por que os bayesianos devem pensar sobre as probabilidades anteriores. Existe um ramo da estatística, às vezes chamado de estatística “ortodoxa” ou “clássica”, que insiste em prestar atenção apenas às verossimilhanças. Mas se você prestar atenção apenas às verossimilhanças, então, eventualmente, alguma hipótese de moeda fixa sempre derrotará a hipótese de moeda justa, um fenômeno conhecido como “sobre-ajuste” da teoria aos dados. Após trinta lançamentos, a verossimilhança é um bilhão de vezes maior para a hipótese de moeda fixa com essa sequência do que para a hipótese de moeda justa.

Somente se a hipótese de moeda fixa (ou melhor, essa hipótese específica de moeda fixa) for um bilhão de vezes menos provável a priori, a hipótese de moeda fixa pode perder para a hipótese de moeda justa.

Se você sacudir a moeda para reiniciá-la e começar a jogar a moeda novamente, e a moeda produzir cara-cara-coroa-coroa-cara:coroa-coroa-coroa-coroa-cara novamente, isso é uma questão diferente. Isso aumenta as chances posteriores da hipótese de moeda fixa para 10:1, mesmo se a probabilidade inicial fosse de apenas 1%.

Da mesma forma, se realizarmos duas medições sucessivas do contador de partículas (ou o preço do supermercado às quartas-feiras) e ambas as medições retornarem 51, a teoria precisa vence por chances de 10:1.

Então, a teoria precisa vence, mas a teoria vaga ainda pontuaria melhor do que nenhuma teoria. Considere uma terceira teoria, a hipótese de distribuição de conhecimento zero ou entropia máxima, que torna igualmente provável qualquer resultado entre zero e 99. Suponha que vejamos o resultado 51. A teoria vaga produziu uma previsão melhor do que a distribuição de entropia máxima — atribuiu uma maior probabilidade ao resultado que observamos. A teoria vaga é, literalmente, melhor do que nada. Suponha que começamos com chances de 1:20 em favor da hipótese de completa ignorância. (Por que chances de 1:20? Existe apenas uma hipótese de completa ignorância e, além disso, é um tipo de hipótese particularmente simples e intuitiva, a Navalha de Occam.) Após ver o resultado de 51, previsto em 9% pela teoria vaga versus 1% por completa ignorância, as chances posteriores vão para 10:20 ou 1:2. Se então virmos outro resultado de 51, as chances posteriores vão para 10:2 ou 83% de probabilidade para a teoria vaga, assumindo não haver teoria mais precisa em consideração.

No entanto, a timidez da teoria vaga — sua relutância em produzir uma previsão exata e aceitar a falsificação em qualquer outro resultado - a torna vulnerável à teoria ousada e precisa. (Contanto, é claro, que a teoria ousada adivinhe corretamente o resultado!) Suponha que as chances anteriores fossem 1:10:200 para as teorias precisa, vaga e ignorante - probabilidades anteriores de 0,5%, 4,7% e 94,8% para as teorias precisa, vaga e ignorante. Esta figura reflete nossa distribuição de probabilidades anteriores sobre classes de hipóteses, com a massa de probabilidade distribuída sobre classes inteiras da seguinte forma: 50% que o fenômeno muda em todos os dígitos, 25% que o fenômeno muda dentro de alguma casa decimal e 25% que o fenômeno repete o mesmo número a cada vez. Uma hipótese de completa ignorância, 10 hipóteses possíveis para uma casa decimal, 100 hipóteses possíveis para um número repetido. Assim, chances anteriores de 1:10:200 para a hipótese precisa 51, a hipótese vaga “cinquenta” e a hipótese de completa ignorância.

Após ver um resultado de 51, com probabilidade atribuída de 90%, 9% e 1%, as chances posteriores vão para 90:90:200 = 9:9:20. Após ver um resultado adicional de 51, as chances posteriores vão para 810:81:20, ou 89%, 9% e 2%. A teoria precisa agora é favorecida em relação à teoria vaga, que por sua vez é favorecida em relação à teoria ignorante.

Agora considere uma teoria estúpida, que prevê uma probabilidade de 90% de ver um resultado entre zero e nove. A teoria estúpida atribui uma probabilidade de 0,1% ao resultado real, 51. Se as chances fossem inicialmente 1:10:200:10 para as teorias precisa, vaga, ignorante e estúpida, as chances posteriores após ver 51 uma vez seriam 90:90:200:1. A teoria estúpida foi falsificada (probabilidade posterior de 0,2%).

É possível ter um modelo tão ruim que seja pior do que nada, se o modelo concentrar sua massa de probabilidade longe do resultado real, fizer previsões confiantes de respostas erradas. Tal hipótese é tão pobre que perde contra a hipótese de completa ignorância. A ignorância é melhor do que o anti-conhecimento.

Observação lateral: No campo da Inteligência Artificial, há uma moda passageira que elogia a glória da aleatoriedade. Ocasionalmente, um pesquisador de IA descobre que, se adicionar ruído a um de seus algoritmos, o algoritmo funciona melhor. Este resultado é relatado com grande entusiasmo, seguido por muitos elogios efusivos aos poderes criativos do caos, imprevisibilidade, espontaneidade, ignorância do que sua própria IA está fazendo, e assim por diante. (Veja The Imagination Engine para um exemplo; de acordo com sua literatura de vendas, eles vendem redes neurais feridas e moribundas. [8]) Mas quão triste é um algoritmo se você pode aumentar seu desempenho injetando entropia em estágios intermediários de processamento? O algoritmo deve ser tão perturbado que parte de seu trabalho se concentra em concentrar a massa de pro-

babilidade longe de boas soluções. Se a injeção de aleatoriedade resultar em uma melhoria confiável, então algum aspecto do algoritmo deve ter um desempenho pior do que aleatório. Somente em IA as pessoas criariam algoritmos literalmente mais burros do que um saco de tijolos, aumentariam os resultados ligeiramente de volta para a ignorância e então argumentariam pelo poder de cura do ruído.

Suponha que em nosso experimento vejamos os resultados 52, 51 e 58. A teoria precisa dá a este evento conjuntivo uma probabilidade de mil para um vezes 90% vezes mil para um, enquanto a teoria mais vaga dá a este evento conjuntivo uma probabilidade de 9% ao cubo, o que dá... oh... hum... vejamos... um milhão para um, dada a teoria precisa, versus mil para um, dada a teoria vaga. Ou por aí; estamos contando potências brutas de dez. Versus um milhão para um, dada a distribuição de conhecimento zero que atribui uma probabilidade igual a todos os resultados. Versus um bilhão para um, dado um modelo pior do que nada, a hipótese estúpida, que afirma uma probabilidade de 90% de ver um número menor que 10. Usando esses números aproximados, a teoria vaga acumula uma pontuação de −30 decibéis (uma probabilidade de 1/1000 para todo o resultado experimental), versus pontuações de −60 para a teoria precisa, −60 para a teoria ignorante e −90 para a teoria estúpida. Nem sempre é verdade que a pontuação mais alta vence, porque precisamos considerar nossas chances anteriores de 1:10:200:10, confidências de −23, −13, 0 e −13 decibéis. A teoria vaga ainda vem com a pontuação total mais alta em -43 decibéis. (Se ignorássemos nossas probabilidades anteriores, cada novo experimento substituiria os resultados acumulados de todos os experimentos anteriores; não poderíamos acumular conhecimento. Além disso, a hipótese de moeda fixa sempre venceria.)

Como sempre, não devemos nos alarmar que mesmo a melhor teoria ainda tenha uma pontuação baixa - lembre-se da parábola da moeda justa. As teorias são aproximações. Em princípio, podemos conseguir prever a sequência exata de lançamentos de moeda. Mas seria preciso tirar medidas melhores e mais poder de computação do que estamos dispostos a gastar. Talvez pudéssemos alcançar uma previsão de 60/40 de lançamentos de moeda, com um modelo bom o suficiente…? Seguimos com a melhor aproximação que temos e tentamos alcançar uma boa calibração, mesmo que a discriminação não seja perfeita.

Conduzimos nossa análise até agora sob as regras da teoria da probabilidade bayesiana, na qual não há como ter mais de 100% de massa de probabilidade e, portanto, nenhuma maneira de trapacear para qualquer resultado poder contar como “confirmação” de sua teoria. Conforme a lei bayesiana, o dinheiro do jogo não pode ser falsificado; você só tem uma quantidade limitada de argila.

Infelizmente, os seres humanos não são bayesianos. Os seres humanos tentam bizarramente defender hipóteses, se esforçando deliberadamente para prová-las ou impedir a refutação. Esse comportamento não tem análogo nas leis da teoria da probabilidade ou da teoria da decisão. Na teoria formal da probabilidade, a hipótese é, e a evidência é, e/ou a hipótese é confirmada, ou não. Na teoria formal da decisão, um agente pode se esforçar para investigar alguma questão da qual o agente está atualmente incerto, não sabendo se a evidência irá para um lado ou para o outro. Em nenhum dos casos se tenta deliberadamente provar uma ideia ou evitar refutá-la. Pode-se testar ideias das quais se está genuinamente incerto, mas não ter um resultado “preferido” da investigação. Não se pode tentar provar hipóteses, nem impedir sua prova. Não consigo transmitir adequadamente o quão ridícula seria a noção para um verdadeiro bayesiano; não há nem mesmo palavras na linguagem de Bayes para descrever o erro…

Para cada expectativa de evidência, há uma expectativa igual e oposta de contra-evidência. Se A é evidência em favor de B, então não-A deve ser evidência em favor de não-B. As forças das evidências podem não ser iguais; evidências raras, mas fortes, em uma direção, podem ser equilibradas por evidências comuns, mas fracas, na outra direção. Mas não é possível que A e não-A sejam evidências em favor de B., ou seja, não é possível sob as leis da teoria da probabilidade.

Os humanos muitas vezes parecem querer ter seu bolo e comê-lo também. Independentemente do resultado que testemunhamos, é aquele que prova nossa teoria. Como Spee, o padre em Conservação da Evidência Esperada, colocou: “O comitê de investigação se sentiria envergonhado se absolvesse uma mulher; uma vez presa e acorrentada, ela tem que ser culpada, por meios justos ou sujos.” [9]

A maneira como a psicologia humana parece funcionar é que primeiro vemos algo acontecer e, em seguida, tentamos argumentar que isso corresponde a qualquer hipótese que tínhamos em mente de antemão. Em vez de massa de probabilidade conservada, para distribuir sobre previsões antecipadas, temos uma

sensação de compatibilidade - o grau em que a explicação e o evento parecem "se ajustar". O "ajuste" não é conservado. Não há equivalente à regra de que a massa de probabilidade deve somar um. Um psicanalista pode explicar qualquer comportamento possível de um paciente construindo uma estrutura apropriada de "racionalizações" e "defesas"; há ajuste, portanto, deve ser verdade.

Agora considere a fábula contada em Explicações Falsas - os alunos vendo um radiador e uma placa de metal ao lado do radiador. Os alunos nunca iriam prever com antecedência que o lado da placa próximo ao radiador seria mais frio. No entanto, vendo o fato, eles conseguiram fazer suas explicações "se ajustarem". Eles perderam sua preciosa chance de perplexidade, para perceber que seus modelos não previam o fenômeno que observaram. Eles sacrificaram sua capacidade de ficar mais confusos com a ficção do que com a verdade. E eles não perceberam que "indução de calor, blá blá blá, portanto, o lado próximo é mais frio" é uma previsão vaga e verbal, espalhada por uma gama enormemente ampla de valores possíveis para temperaturas medidas específicas. A aplicação de equações de difusão e equilíbrio daria uma previsão precisa para possíveis valores conjuntos. Pode não especificar os primeiros valores que você mediu, mas quando você soubesse alguns valores, poderia gerar uma previsão precisa para o resto. A pontuação para todo o resultado experimental seria muito melhor do que qualquer alternativa menos precisa, especialmente uma previsão vaga e verbal.

Agora você tem uma explicação técnica da diferença entre uma explicação verbal e uma explicação técnica. É uma explicação técnica porque permite calcular exatamente o quão técnica é uma explicação. Hipóteses vagas podem ser tão vagas que apenas uma inteligência sobre-humana poderia calcular exatamente o quão vagas. Talvez uma inteligência suficientemente grande pudesse extrapolar todos os resultados experimentais possíveis e extrapolar todos os veredictos possíveis do adivinhador vago sobre o quão bem a hipótese vaga "se encaixa" e então renormalizar a distribuição de "ajuste" em uma distribuição de verossimilhança que somasse um. Mas, em princípio, ainda se pode calcular exatamente o quão vaga é uma hipótese vaga. O cálculo simplesmente não é computacionalmente tratável, da mesma forma que calcular trajetórias de avião por meio da mecânica quântica não é computacionalmente tratável.

Sustento que todos precisam aprender pelo menos um assunto técnico: física, ciência da computação, biologia evolutiva, teoria da probabilidade bayesiana ou algo assim. Alguém sem nenhum assunto técnico em seu currículo não tem referência para o que significa "explicar" algo. Eles podem pensar que "Tudo é Fogo" é uma explicação, como fez o filósofo grego Heráclito. Portanto, defendo que a teoria da probabilidade bayesiana seja ensinada no ensino médio. A teoria da probabilidade bayesiana é a única parte da matemática que conheço que é acessível no nível do ensino médio e que permite uma compreensão técnica de um assunto - a dinâmica da crença - o qual é um domínio real do dia a dia e tem consequências emocionalmente significativas. Estudar a probabilidade bayesiana daria aos alunos uma referência para o que significa "explicar" algo.

Muitos acadêmicos pensam que ser "técnico" significa falar em polissílabos secos. Aqui está uma explicação "técnica" da explicação técnica:

As equações da teoria da probabilidade favorecem hipóteses que preveem fortemente os dados observados exatos. Modelos fortes concentram ousadamente sua densidade de probabilidade em resultados precisos, tornando-os falsificáveis se os dados atingirem outro lugar e dando-lhes tremendas vantagens de verossimilhança sobre modelos menos ousados, menos precisos. A explicação verbal é executada na avaliação psicológica da compatibilidade post facto não conservada em vez da densidade de probabilidade ante facto conservada. E a explicação verbal não pinta imagens nitidamente detalhadas, implicando uma distribuição de verossimilhança suave nas proximidades dos dados.

Isso é satisfatório? Não. Ouça as frases impressionantes e pesadas, ressoando com o baque surdo da perícia. Veja os infelizes alunos, escrevendo essas frases em uma folha de papel. Mesmo depois que os ouvintes ouvem as palavras rituais, eles não podem realizar cálculos. Você conhece a matemática, então as palavras são significativas. Você pode realizar os cálculos após ouvir as palavras impressionantes, assim como poderia ter feito antes. Mas e aquele que não viu nenhum cálculo realizado? Que novas habilidades eles ganharam com aquela palestra "técnica", exceto a habilidade de recitar palavras fascinantes?

"Bayesiano" com certeza é uma palavra fascinante, não é? Vamos tirá-la de nossos sistemas: Bayes

Bayes Bayes Bayes Bayes Bayes Bayes Bayes Bayes...

A sílaba sagrada não tem sentido, exceto enquanto diz a alguém para aplicar matemática. Portanto, quem ouve já deve conhecer a matemática.

Por outro lado, se você conhece a matemática, pode ser tão bobo quanto quiser e ainda técnico.

Assim, dispomos de mais um estereótipo de racionalidade, de que a racionalidade consiste em formalidade árida e solenidade sem humor. O que isso tem a ver com o problema de distinguir a verdade da falsidade? O que isso tem a ver com obter o mapa que reflete o território? Um cientista digno de um jaleco de laboratório deve conseguir fazer descobertas originais enquanto usa um traje de palhaço, ou dar uma palestra com uma voz estridente de inalar hélio. Não está escrito em lugar nenhum na matemática da teoria da probabilidade que não se pode se divertir. A lâmina que corta até a resposta correta não tem dignidade ou tolice em si, embora caiba na mão de um portador tolo.

Um modelo útil não é apenas algo que você sabe, como você sabe que um avião é feito de átomos. Um modelo útil é o conhecimento que você pode calcular em tempo razoável para prever eventos reais que você sabe como observar. Talvez alguém descubra que, usando um modelo que viola a Conservação do Momento apenas um pouco, você pode calcular a aerodinâmica do 747 muito mais barato do que se insistir que o momento é exatamente conservado. Então, se você tem dois computadores competindo para produzir a melhor previsão, pode ser que a melhor previsão venha do modelo que viola a Conservação do Momento. Isso não significa que o 747 viola a Conservação do Momentum na realidade. Nenhum dos modelos usa átomos individuais, mas isso não implica que o 747 não seja feito de átomos. Os físicos usam modelos diferentes para prever aviões e colisões de partículas porque seria muito caro calcular o avião partícula por partícula.

Você provaria que o 747 é feito de átomos com dados experimentais que os modelos aerodinâmicos não conseguiam lidar; por exemplo, você treinaria um microscópio de tunelamento de varredura em uma seção da asa e olharia para os átomos. Da mesma forma, você pode usar um instrumento de medição mais fino para discriminar entre um 747 que realmente desobedeceu à Conservação do Momento como a aproximação barata previu, versus um 747 que obedeceu à Conservação do Momento como a física subjacente previu. A teoria vencedora é aquela que melhor prevê todas as previsões experimentais juntas.

Nossa regra de pontuação bayesiana nos dá uma maneira de combinar os resultados de todos os nossos experimentos, mesmo experimentos que usam métodos diferentes. Além disso, a teoria atômica permite, abraça e, em certo sentido, exige o modelo aerodinâmico. Ao pensar abstratamente sobre as suposições da teoria atômica, percebemos que o modelo aerodinâmico deve ser uma aproximação boa (e muito mais barata) da teoria atômica, e então a teoria atômica apoia o modelo aerodinâmico, em vez de competir com ele. Uma teoria bem-sucedida pode abranger muitos modelos para diferentes domínios, contanto que os modelos sejam reconhecidos como aproximações e, em cada caso, o modelo seja compatível com (ou idealmente exigido pela) teoria subjacente.

Nossa física fundamental - a mecânica quântica, a família padrão de partículas e a relatividade - é uma teoria que abrange uma enorme família de modelos para fenômenos físicos macroscópicos. Existe a física dos líquidos, dos sólidos e dos gases; no entanto, isso não significa que existam coisas fundamentais no mundo que tenham a propriedade intrínseca de liquidez.

> Aparentemente, há cor, aparentemente doçura, aparentemente amargor, na verdade, existem apenas átomos e o vazio[53].
>
> — Demócrito, 420 a.C., de Robinson e Groves [10]

Ao argumentar que uma teoria "técnica" deve ser definida como uma teoria que concentra fortemente a probabilidade em previsões antecipadas específicas, estou estabelecendo um padrão de rigor ex-

53 NT. Texto original em inglês. *Apparently there is colour, apparently sweetness, apparently bitterness, actually there are only atoms and the void.*

tremamente alto. Vimos que uma teoria vaga pode ser melhor do que nada. Uma teoria vaga pode vencer a hipótese da ignorância, se não houver teorias precisas para competir contra ela.

Existe uma enorme família de modelos pertencentes à teoria central subjacente da vida e da biologia, a teoria subjacente que às vezes é chamada de neodarwinismo, seleção natural ou evolução. Alguns modelos na teoria evolutiva são quantitativos. A maneira como o DNA codifica proteínas é redundante; duas sequências de DNA diferentes podem codificar a mesma proteína. Existem quatro bases de DNA {A,T,C,G} e 64 combinações possíveis de três bases de DNA. Mas esses 64 códons possíveis descrevem apenas 20 aminoácidos mais um código de parada. A deriva genética deve, portanto, produzir mudanças não funcionais nos genomas das espécies, por meio de mutações que por acaso se fixam no pool genético. A taxa de acumulação de diferenças não funcionais entre os genomas de duas espécies com um ancestral comum depende de parâmetros como o número de gerações decorridas e a intensidade da seleção naquele locus genético. Esse é um exemplo de um membro da família de modelos evolutivos que produz previsões quantitativas. Existem também frequências alélicas de desequilíbrio sob seleção, equilíbrios estáveis para estratégias de teoria dos jogos, proporções sexuais, etc.

Tudo isso está sob o título de "palavras fascinantes". Infelizmente, existem certas facções religiosas que espalham desinformação grosseira sobre a teoria evolutiva. Portanto, enfatizo que muitos modelos na teoria evolutiva fazem previsões quantitativas confirmadas experimentalmente, e que tais modelos são muito mais do que suficientes para demonstrar que, por exemplo, humanos e chimpanzés são relacionados por um ancestral comum. Se você foi vítima de desinformação criacionista - ou seja, se você ouviu alguma sugestão de que a teoria evolutiva é controversa ou intestável ou "apenas uma teoria" ou não rigorosa ou não técnica, ou de alguma forma não confirmada por uma quantidade inimaginavelmente enorme de evidências experimentais. Recomendo ler o FAQ do TalkOrigins [11] e estudar biologia evolutiva com matemática.

Mas imagine voltar no tempo para o século XIX, quando a teoria da seleção natural havia acabado de ser descoberta por Charles Darwin e Alfred Russel Wallace. Imagine o evolucionismo logo após seu nascimento, quando a teoria não tinha nada remotamente parecido com o corpo moderno de modelos quantitativos e grandes montanhas de evidências experimentais. Não havia como saber que se descobriria que os seres humanos e os chimpanzés tinham 95% de material genético compartilhado. Ninguém sabia que o DNA existia. Mesmo assim, os cientistas se aglomeraram na nova teoria da seleção natural. E mais tarde descobriu-se que havia um material genético copiado com precisão com o potencial de sofrer mutação, que humanos e chimpanzés eram comprovadamente relacionados, etc.

Portanto, o padrão muito estrito e muito alto que propus para uma teoria "técnica" é muito estrito. Historicamente, tem sido possível discriminar com sucesso teorias verdadeiras de teorias falsas, com base em previsões do tipo que chamei de "vagas". Previsões vagas de, digamos, 80% de confiança, podem construir uma grande vantagem sobre hipóteses alternativas, dados experimentos suficientes. Talvez uma teoria desse tipo, produzindo previsões que não são precisamente detalhadas, mas que, no entanto, estão corretas, poderia ser chamada de "semi técnica"?

Mas certamente as teorias técnicas são mais confiáveis do que as teorias semi técnicas? Certamente as teorias técnicas devem ter precedência, comandar maior respeito? Certamente a física, que produz previsões extremamente exatas, é em certo sentido melhor confirmada do que a teoria evolutiva? Não estou insinuando que a teoria evolutiva está errada, é claro; mas por mais vastas que sejam as montanhas de evidências que favorecem a evolução, a física não vai além por meio de vastas montanhas de confirmação experimental precisa? Observações de estrelas de nêutrons confirmam as previsões da Relatividade Geral com precisão de uma parte em cem trilhões (10^{14}). O que a teoria evolutiva tem para igualar isso?

Daniel Dennett disse certa vez que, medido pela simplicidade da teoria e pela quantidade de complexidade que ela explicava, Darwin teve a maior ideia da história do tempo. [12]

Certa vez, houve um conflito entre a física do século XIX e o evolucionismo do século XIX. De acordo com os melhores modelos físicos então em uso, o Sol não poderia estar queimando por muito tempo. Três mil anos com energia química ou 40 milhões de anos com energia gravitacional. Não havia fonte de energia conhecida pela física do século XIX que permitisse uma queima mais longa. A física do século XIX não era tão poderosa quanto a física moderna - não tinha previsões precisas com precisão de uma parte em 10^{14}. Mas a física do

século XIX ainda tinha o caráter matemático da física moderna, uma disciplina cujos modelos produziam previsões detalhadas, precisas e quantitativas. A teoria evolutiva do século XIX era totalmente semi técnica, sem um pingo de modelagem quantitativa. Nem mesmo os experimentos de Mendel com ervilhas eram conhecidos na época. No entanto, parecia provável que a evolução exigiria mais do que insignificantes 40 milhões de anos para operar - centenas de milhões, até bilhões de anos. A antiguidade da Terra era uma previsão vaga e semi técnica, de uma teoria vaga e semi técnica. Em contraste, os físicos do século XIX tinham um modelo preciso e quantitativo, que por meio de cálculo formal produziu o ditado preciso e quantitativo de que o Sol simplesmente não poderia ter queimado por tanto tempo.

> As limitações dos períodos geológicos, impostas pela ciência física, não podem, é claro, refutar a hipótese da transmutação de espécies; mas parece suficiente para refutar a doutrina de que a transmutação ocorreu por meio da "descendência com modificação por seleção natural".[54]
>
> — Lord Kelvin, de Lyle Zapato [13]

A história registra quem venceu.

A moral? Se você pode dar previsões antecipadas com 80% de confiança em perguntas sim ou não, pode ser uma teoria "vaga"; pode estar errado uma vez em cinco; mas você ainda pode construir uma grande vantagem de pontuação sobre a hipótese da ignorância. O suficiente para confirmar uma teoria, se não houver melhores concorrentes. A realidade é consistente; toda teoria correta sobre o universo é compatível com todas as outras teorias corretas. Mapas imperfeitos podem entrar em conflito, mas há apenas um território. O evolucionismo do século XIX pode ter sido uma disciplina semi técnica, mas ainda estava correto (como sabemos agora) e de longe a melhor explicação (mesmo naquela época). Qualquer conflito entre o evolucionismo e outra teoria bem confirmada tinha que refletir algum tipo de anomalia, um erro na afirmação de que as duas teorias eram incompatíveis. A física do século XIX não conseguia modelar a dinâmica do Sol - eles não sabiam sobre reações nucleares. Eles não podiam mostrar que sua compreensão do Sol estava correta em detalhes técnicos, nem calcular a partir de um modelo confirmado do Sol para determinar há quanto tempo o Sol existia. Então, em retrospecto, podemos dizer algo como: "Havia espaço para a possibilidade de que a física do século XIX simplesmente não entendesse o Sol."

Mas isso é retrospectiva. A verdadeira lição é que, embora a física do século XIX fosse precisa e quantitativa, ela não dominava automaticamente a teoria semi técnica do evolucionismo do século XIX. As teorias eram ambas bem apoiadas. Ambas estavam corretas nos domínios sobre os quais foram generalizadas. O aparente conflito entre eles era uma anomalia, e a anomalia acabou decorrendo da incompletude e aplicação incorreta da física do século XIX, não da incompletude e aplicação incorreta do evolucionismo do século XIX. Mas seria inútil comparar a montanha de evidências que apoiam uma teoria versus a montanha de evidências que apoiam a outra. Mesmo naquela época, ambas as montanhas eram grandes demais para supor que qualquer uma das teorias estava simplesmente errada. Montanhas de evidências tão grandes não podem ser colocadas para competir, como se uma falsificasse a outra. Você deve estar aplicando uma teoria incorretamente ou aplicando um modelo fora do domínio que ele prevê bem.

Portanto, você não deve zombar necessariamente de uma teoria apenas porque ela é semi técnica. As teorias semi técnicas podem construir pontuações altas o suficiente, em comparação com todas as alternativas disponíveis, para você saber que a teoria está pelo menos aproximadamente correta. Algum dia, a teoria semi técnica pode ser substituída ou até falsificada por um concorrente mais preciso, mas isso também é verdade para as teorias técnicas. Pense em como a Relatividade Geral de Einstein devorou a teoria da gravitação de Newton.

Mas a correção de uma teoria semi técnica - uma teoria que atualmente não possui modelos precisos e computacionalmente tratáveis, testáveis por experimentos viáveis - pode ser muito menos clara do que a correção de uma teoria técnica. É preciso habilidade, paciência e exame para distinguir boas teorias semi técnicas de teorias simplesmente confusas. Isso não é nada que os humanos fazem bem por instinto, e é por

54 NT. Texto original em inglês. *The limitations of geological periods, imposed by physical science, cannot, of course, disprove the hypothesis of transmutation of species; but it does seem sufficient to disprove the doctrine that transmutation has taken place through "descent with modification by natural selection."*

isso que temos a Ciência.

As pessoas se precipitam e se aproveitam de qualquer razão disponível para rejeitar uma teoria que não gostam. É por isso que dei o exemplo do evolucionismo do século XIX, para mostrar por que não se deve ser muito rápido em rejeitar uma teoria "não técnica" imediatamente. Pelos costumes morais da ciência, o evolucionismo do século XIX era culpado de mais de um pecado. O evolucionismo do século XIX não fez previsões quantitativas. Não estava prontamente sujeito à falsificação. Era na maioria uma explicação do que já havia sido visto. Faltava um mecanismo subjacente, pois ninguém na época sabia sobre DNA. Ele até contradizia as leis da física do século XIX. No entanto, a seleção natural era uma explicação post-facto tão incrivelmente boa que as pessoas se aglomeraram nela, e elas estavam certas. A Ciência, como um esforço humano, requer previsão antecipada. A teoria da probabilidade, como matemática, não distingue entre previsão post-facto e antecipada, porque a teoria da probabilidade assume que as distribuições de probabilidade são propriedades fixas de uma hipótese.

A regra sobre a previsão antecipada é uma regra do processo social da ciência - um costume moral e não um teorema. O costume moral existe para evitar que os seres humanos cometam erros humanos que são difíceis até mesmo de descrever na linguagem da teoria da probabilidade, como mexer depois do fato com o que você afirma que sua hipótese prevê. As pessoas concluíram que o evolucionismo do século XIX era uma excelente explicação, mesmo que fosse post-facto. Esse raciocínio estava correto como teoria da probabilidade, e é por isso que funcionou apesar de todos os pecados científicos. A teoria da probabilidade é matemática. O processo social da ciência é um conjunto de convenções legais para impedir que as pessoas trapaceiem na matemática.

No entanto, também é verdade que, em comparação com um teórico evolucionista moderno, os teóricos evolucionistas do final do século XIX e início do século XX muitas vezes se perdiam tristemente. Darwin, que era brilhante o suficiente para inventar a teoria, acertou uma quantidade incrível. Mas os sucessores de Darwin, que eram apenas brilhantes o suficiente para aceitar a teoria, entenderam mal a evolução com frequência e seriedade. O processo usual da ciência era então necessário para corrigir seus erros. É incrível como poucos erros de raciocínio Darwin [14] cometeu em A Origem das Espécies e A Descendência do Homem, em comparação com aqueles que o seguiram.

Esse também é um risco de uma teoria semi técnica. Mesmo depois que o lampejo de insight genial é confirmado, cientistas meramente médios podem falhar em aplicar os insights adequadamente na ausência de modelos formais. Até a década de 1960, os biólogos falavam da evolução trabalhando "para o bem da espécie" ou sugeriam que os indivíduos restringiriam sua reprodução para evitar a superpopulação de espécies em um habitat. Os melhores teóricos evolucionistas sabiam melhor, mas os teóricos médios não. [15]

Portanto, é muito melhor ter uma teoria técnica do que uma teoria semi técnica. Infelizmente, a Natureza nem sempre é tão gentil a ponto de se tornar descritível por modelos organizados, formais e computacionalmente tratáveis, nem sempre fornece aos Seus alunos instrumentos de medição que podem sondar diretamente Seus fenômenos. Às vezes é apenas uma questão de tempo. O evolucionismo do século XIX era semi técnico, mas depois veio a matemática da genética populacional e, eventualmente, o sequenciamento de DNA. A Natureza nem sempre lhe dará um fenômeno que você pode descrever com modelos técnicos quinze segundos após ter o insight básico.

No entanto, a vanguarda da ciência, a controvérsia, é mais frequentemente sobre uma teoria semi técnica ou um absurdo se passando por uma teoria semi técnica. No momento em que uma teoria atinge o status técnico, ela geralmente não é mais controversa (entre os cientistas). Portanto, a questão de como distinguir boas teorias semi técnicas de absurdos é muito importante para os cientistas, e não é tão fácil quanto descartar imediatamente qualquer teoria que não seja técnica. Para o fim de distinguir a verdade da falsidade, existe toda a disciplina da racionalidade. A arte não é redutível a uma lista de verificação, ou pelo menos, nenhuma lista de verificação que um cientista médio possa aplicar de forma confiável após uma hora de treinamento.Se fosse tão simples, não precisaríamos da ciência.

Por que você presta atenção às controvérsias científicas? Por que pastar em alimentos tão escassos e podres quanto a mídia oferece, quando há tantas refeições sólidas para serem encontradas em livros didáticos? A ciência do livro didático é linda! A ciência do livro didático é compreensível, ao contrário de meras

palavras fascinantes que nunca podem ser verdadeiramente belas. Palavras fascinantes não têm poder, nem mesmo qualquer significado, sem a matemática. As palavras fascinantes não são conhecimentos, mas a ilusão de conhecimento, e é por isso que traz tão pouca satisfação saber que "a gravidade resulta da curvatura do espaço-tempo". A ciência não está nas palavras fascinantes, embora seja tudo o que você lerá como notícias de última hora.

Pode haver justificativa para seguir uma controvérsia científica. Você pode ser um especialista nesse campo, caso em que essa controvérsia científica é sua carne adequada. Ou a controvérsia científica pode ser algo que você precisa saber agora, porque afeta sua vida. Talvez seja o século XIX, e você está olhando lascivamente para um membro do sexo apropriado usando um traje de banho do século XIX, e você precisa saber se seu desejo sexual vem de uma psicologia construída pela seleção natural, ou é uma tentação colocada em você pelo Diabo para atraí-lo para o fogo do inferno.

Não é totalmente impossível que encontremos uma controvérsia científica que nos afete e descubramos que temos uma necessidade urgente e ardente da resposta correta. Portanto, discutirei alguns dos sinais de alerta que historicamente distinguiram hipóteses vagas que mais tarde se revelaram besteiras não científicas de hipóteses vagas que mais tarde se tornaram teorias confirmadas. Apenas lembre-se da lição histórica do evolucionismo do século XIX e resista à tentação de reprovar todas as teorias que perdem um único item em sua lista de verificação. Não é minha intenção dar às pessoas outra desculpa para descartar a boa ciência que as incomoda. Se você aplicar critérios mais rígidos às teorias de que não gosta do que às teorias de que gosta (ou vice-versa!), então cada novo detalhe que você aprende a escolher, cada nova falha lógica que você aprende a detectar, o torna muito mais estúpido. A inteligência, para ser útil, deve ser usada para algo além de se derrotar.

Um dos sinais clássicos de uma hipótese ruim é que ela deve despender grande esforço para evitar a falsificação - elaborando razões pelas quais a hipótese é compatível com o fenômeno, embora o fenômeno não tenha se comportado como esperado. Carl Sagan dá o exemplo de alguém que afirma que um dragão vive em sua garagem. Sagan originalmente tirou a lição de que hipóteses ruins precisam fazer um trabalho rápido para evitar a falsificação - para manter uma aparência de "ajuste". [16]

Gostaria de salientar que o reclamante obviamente tem um bom modelo da situação em algum lugar em sua cabeça, porque ele pode prever, com antecedência, exatamente quais desculpas ele vai precisar. Para um bayesiano, uma hipótese não é algo que você afirma em voz alta e enfática. Uma hipótese é algo que controla suas antecipações, as probabilidades que você atribui a experiências futuras. Isso é o que é uma probabilidade para um bayesiano - é isso que você pontua, é isso que você calibra. Então, embora nosso reclamante possa dizer em voz alta, enfaticamente e honestamente que acredita que há um dragão invisível na garagem, ele não antecipa que há um dragão invisível na garagem - ele antecipa exatamente a mesma experiência que o cético.

Quando julgo as previsões de uma hipótese, pergunto quais experiências eu anteciparia, não em quais fatos eu acreditaria.

O outro lado:

Recentemente, discuti com um amigo meu sobre uma questão da teoria evolutiva. Meu amigo alegou que o agrupamento de mudanças no registro fóssil (aparentemente, existem períodos de estase comparativa seguidos por mudanças comparativamente acentuadas; em si uma observação controversa conhecida como "equilíbrio pontuado") mostrou que havia algo errado com nossa compreensão da especiação. Meu amigo pensou haver alguma força desconhecida em ação - não sobrenatural, mas alguma consideração natural que a teoria evolutiva padrão não considerava. Como meu amigo não deu uma hipótese concorrente específica que produzisse melhores previsões, sua tese tinha que ser que o modelo evolutivo padrão era estúpido em relação aos dados - que o modelo padrão fez uma previsão específica que estava errada; que o modelo se saiu pior do que a ignorância completa ou algum outro concorrente padrão.

No início, caí na armadilha; aceitei a suposição implícita de que o modelo padrão previa suavidade e baseei meu argumento em minha lembrança de que as mudanças no registro fóssil não eram tão acentuadas quanto ele afirmava. Ele me desafiou a produzir um intermediário evolutivo entre o *Homo erectus* e o *Homo*

sapiens; pesquisei no Google e encontrei o *Homo heidelbergensis*. Ele me parabenizou e reconheceu que eu havia marcado um ponto importante, mas ainda insistiu que as mudanças eram muito abruptas e não constantes o suficiente. Comecei a explicar por que pensei que um padrão de mudança desigual poderia surgir do modelo padrão: as pressões de seleção ambiental podem não ser constantes… "Aha!" meu amigo disse, "você está inventando suas desculpas com antecedência."

Mas suponha que o registro fóssil mostrasse, em vez disso, um conjunto de mudanças suaves e graduais. Meu amigo poderia ter argumentado que o modelo padrão de evolução como um processo caótico e ruidoso não poderia explicar essa suavidade? Se é um pecado científico afirmar post facto que nossa amada hipótese prevê os dados, não deveria ser igualmente um pecado afirmar post facto que a hipótese concorrente é estúpida com os dados?

Se uma hipótese tem um modelo puramente técnico, não há problema; podemos calcular a previsão do modelo formalmente, sem variáveis informais para fornecer uma alça para intromissão post facto. Mas e as teorias semi técnicas? Obviamente, uma teoria semi técnica deve produzir algumas boas previsões antecipadas sobre algo, ou então por que se preocupar? Mas depois que a teoria é semi-confirmada, os detratores podem alegar que os dados mostram um problema com a teoria semi técnica, quando o "problema" é construído post facto? No mínimo, os detratores devem ser muito específicos sobre quais dados um modelo confirmado prevê estupidamente e por que o modelo confirmado deve fazer (post -facto) essa previsão estúpida. Quão acentuada é uma mudança "muito acentuada", quantitativamente, para o modelo padrão de evolução permitir? Exatamente quanta estabilidade você acha que o modelo padrão de evolução prevê? Como você sabe? É tarde demais para dizer isso, depois que você viu os dados?

Quando meu amigo me acusou de inventar desculpas, parei e me perguntei de quais desculpas eu antecipava precisar fazer. Decidi que minha compreensão atual da teoria evolutiva não dizia nada sobre se a taxa de mudança evolutiva deveria ser intermitente e irregular, ou suave e gradual. Se eu não tivesse visto o gráfico com antecedência, não poderia tê-lo previsto. (Infelizmente, eu proferi esse veredicto após ver os dados…) Talvez existam modelos na família evolutiva que fariam previsões antecipadas de estabilidade ou variabilidade, mas se houverem, eu não sei sobre eles. Mais precisamente, meu amigo também não sabia.

Nem sempre é sábio perguntar aos oponentes de uma teoria o que seus concorrentes preveem. Obtenha as previsões da teoria dos melhores defensores da teoria. Apenas se certifique de anotar suas previsões com antecedência. Sim, às vezes os defensores de uma teoria tentam fazer a teoria "se ajustar" a evidências que claramente não se encaixam. Mas se você se encontrar se perguntando o que uma teoria prevê, pergunte primeiro entre os defensores da teoria e depois peça aos detratores para fazerem o interrogatório.

Além disso: os modelos podem incluir ruído. Se levantarmos a hipótese de que os dados estão tendendo para cima lenta e continuamente, mas nosso instrumento de medição tem um erro de 5%, então não adianta apontar para um ponto de dados que mergulha abaixo do ponto de dados anterior e gritar triunfantemente: "Veja! Ele desceu! Para baixo, para baixo, para baixo! E não me diga por que sua teoria se encaixa no mergulho; você está apenas inventando desculpas! " Modelos formais e técnicos geralmente incorporam termos de erro explícitos. O termo de erro espalha a densidade de verossimilhança, diminui a precisão do modelo e reduz a pontuação da teoria, mas a regra de pontuação bayesiana ainda governa. Um modelo técnico pode permitir erros e cometer erros e ainda se sair melhor do que a ignorância. Em nosso exemplo de supermercado, mesmo a hipótese precisa de 51 ainda aposta apenas 90% de sua massa de probabilidade em 51; a hipótese precisa afirma apenas que 51 acontece nove vezes em dez. Ignorar nove 51s, apontar para um caso de 82 e cantar em triunfo não constitui uma refutação. Isso não é uma desculpa, é uma previsão antecipada explícita de um modelo técnico.

O termo de erro torna a teoria "precisa" vulnerável a uma alternativa superprecisa que previu o 82. O modelo padrão também seria vulnerável a um modelo precisamente ignorante que previu uma chance de 60% de 51 na rodada em que vimos 82, espalhando a verossimilhança mais entropicamente naquele erro particular. Não importa quão boa seja a teoria, a ciência sempre tem espaço para um competidor com pontuação mais alta. Mas se você não apresentar uma alternativa melhor, se tentar apenas mostrar que uma teoria aceita é estúpida em relação aos dados, esse esforço científico pode ser mais exigente do que apenas substituir a teoria antiga por uma nova.

Os astrônomos registraram o avanço inexplicável do periélio de Mercúrio, não explicado pela física newtoniana - ou melhor, a física newtoniana previu 5.557 segundos de arco por século, onde a quantidade observada foi 5.600. [17] Mas os cientistas daquela época deveriam ter descartado a gravitação newtoniana com base em tão pequenas contra-evidências inexplicáveis? O que eles teriam usado em vez disso? Eventualmente, a teoria da gravitação de Newton foi posta de lado, depois que a Relatividade Geral de Einstein explicou precisamente a discrepância orbital de Mercúrio e também fez previsões antecipadas bem-sucedidas. Mas não havia como saber de antemão que as coisas seriam assim.

No século XIX, havia uma anomalia persistente na órbita de Urano. As pessoas diziam: "Talvez a lei de Newton comece a falhar em longas distâncias." Eventualmente, alguns companheiros brilhantes olharam para a anomalia e disseram: "Isso poderia ser um planeta externo desconhecido?" Urbain Le Verrier e John Couch Adams, independentemente, fizeram alguns rabiscos e cálculos, usando a teoria padrão de Newton - e previram a localização de Netuno com precisão de um grau de arco, confirmando dramaticamente a gravitação newtoniana. [18]

Somente depois que a Relatividade Geral produziu precisamente o avanço do periélio de Mercúrio, soubemos que a gravitação newtoniana nunca o explicaria.

Na Explicação Intuitiva, vimos como o insight de Karl Popper de que a falsificação é mais forte do que a confirmação se traduz em uma verdade bayesiana sobre as razões de verossimilhança. Popper errou ao pensar que a falsificação era qualitativamente diferente da confirmação; ambos são regidos pelas mesmas regras bayesianas. Mas a filosofia de Popper refletia uma verdade importante sobre uma diferença quantitativa entre falsificação e confirmação.

Popper ficou profundamente impressionado com as diferenças entre as teorias supostamente "científicas" de Freud e Adler e a revolução efetuada pela teoria da relatividade de Einstein na física nas primeiras duas décadas deste século. A principal diferença entre eles, como Popper viu, era que, embora a teoria de Einstein fosse altamente "arriscada", no sentido de que era possível deduzir consequências dela que eram, à luz da então dominante física newtoniana, altamente improváveis (por exemplo, que a luz é desviada em direção a corpos sólidos - confirmado pelos experimentos de Eddington em 1919), e que, se se revelassem falsos, falsificariam toda a teoria, nada poderia, mesmo em princípio, falsificar as teorias psicanalíticas. Essas últimas, Popper concluiu, têm mais em comum com mitos primitivos do que com ciência genuína. Isso quer dizer, ele viu que o que aparentemente é a principal fonte de força da psicanálise, e a principal base sobre a qual se baseia sua reivindicação de status científico, ou seja, sua capacidade de acomodar e explicar todas as formas possíveis de comportamento humano, é na verdade uma fraqueza crítica, pois acarreta que não é, e não poderia ser, genuinamente preditiva. As teorias psicanalíticas, por sua natureza, são insuficientemente precisas para ter implicações negativas e, portanto, estão imunizadas contra a falsificação experiencial…

Popper, então, repudia a indução e rejeita a visão de que é o método característico de investigação e inferência científica, e substitui a falseabilidade em seu lugar. É fácil, ele argumenta, obter evidências em favor de virtualmente qualquer teoria, e ele consequentemente sustenta que tal "corroboração", como ele a chama, deve contar cientificamente apenas se for o resultado positivo de uma previsão genuinamente "arriscada", que poderia concebivelmente ter sido falsa. Para Popper, uma teoria é científica apenas se for refutável por um evento concebível. Todo teste genuíno de uma teoria científica, então, é logicamente uma tentativa de refutá-la ou falsificá-la…

Toda teoria científica genuína, então, na visão de Popper, é proibitiva, no sentido de que proíbe, por implicação, eventos ou ocorrências particulares. [19]

Na filosofia de Popper, a força de uma teoria científica não é o quanto ela explica, mas o quanto ela não explica. A virtude de uma teoria científica não está nos resultados que ela permite, mas nos resultados que ela proíbe. As teorias de Freud, que pareciam explicar tudo, não proibiam nada.

Traduzindo isso em termos bayesianos, descobrimos que quanto mais resultados um modelo proíbe, mais densidade de probabilidade o modelo concentra nos resultados restantes permitidos. Quanto mais resultados uma teoria proíbe, maior o conteúdo de conhecimento da teoria. Quanto mais ousadamente uma teoria se expõe à falsificação, mais definitivamente ela lhe diz quais experiências antecipar.

Uma teoria que pode explicar qualquer experiência corresponde a uma hipótese de completa ignorância - uma distribuição uniforme com densidade de probabilidade espalhada uniformemente sobre todos os resultados possíveis.

O flogisto era a resposta do século XVIII ao Fogo Elemental dos alquimistas gregos. Você não poderia usar a teoria do flogisto para prever o resultado de uma transformação química - primeiro você olhava para o resultado, depois usava o flogisto para explicá-lo. A teoria do flogisto era infinitamente flexível; uma hipótese disfarçada de conhecimento zero. Da mesma forma, a teoria do vitalismo não explica como a mão se move, nem lhe diz quais transformações esperar da química orgânica; e o vitalismo certamente não permite cálculos quantitativos.

O outro lado:

Cuidado com o pensamento de lista de verificação: ter um mistério sagrado, ou uma resposta misteriosa, não é o mesmo que se recusar a explicar algo. Alguns elementos em nossa física são considerados "fundamentais", ainda não reduzidos ou explicados. Mas esses elementos fundamentais de nossa física são regidos por regras causais claramente definidas, matematicamente simples e formalmente computáveis.

Ocasionalmente, algum maluco se opõe à física moderna alegando que ela não fornece um "mecanismo subjacente" para uma lei matemática atualmente tratada como fundamental. (Afirmar que uma lei matemática carece de um "mecanismo subjacente" é uma das entradas no "Índice de Maluco" de John Baez. [20]) O "mecanismo subjacente" que o maluco propõe em resposta é vago, verbal e não produz aumento no poder preditivo - caso contrário, não classificaríamos o reclamante como um maluco.

Nossa física atual torna o campo eletromagnético fundamental e se recusa a explicá-lo mais. Mas o "campo eletromagnético" é um fundamental regido por regras matemáticas claras, sem propriedades fora das regras matemáticas, sujeito a computação formal para descrever seu efeito causal sobre o mundo. Algum dia, alguém pode sugerir uma matemática aprimorada que produza melhores previsões, mas eu não indiciaria o modelo atual por motivos de mistério. Uma teoria que inclui elementos fundamentais não é o mesmo que uma teoria que contém elementos misteriosos.

Os fundamentos devem ser simples. "Vida" não é um bom fundamental, "oxigênio" é um bom fundamental e "campo eletromagnético" é um fundamental melhor. A vida pode parecer simples para um vitalista - é a capacidade simples e mágica de seus músculos se moverem sob sua direção mental. Por que a vida não deveria ser explicada por uma substância fundamental simples e mágica como elã vital? Mas fenômenos que parecem psicologicamente muito simples - pequenos pontos de luz no céu, chama laranja-brilhante quente, carne se movendo sob direção mental - muitas vezes escondem vastas profundezas de complexidade subjacente. A proposição de que a vida é um fenômeno complexo pode parecer incrível para o vitalista, olhando para um mistério opaco e vazio, sem alças óbvias; mas sim, Virgínia, existe uma complexidade subjacente. O critério de simplicidade relevante para a Navalha de Ocam é a simplicidade matemática ou computacional. Uma vez que reduzimos nosso modelo a elementos fundamentais matematicamente simples, não compartilhando em si as qualidades misteriosas do mistério, interagindo de maneiras claramente definidas para produzir o fenômeno anteriormente misterioso como uma previsão detalhada, isso é tão não misterioso quanto a humanidade já descobriu como fazer qualquer coisa.

Muitas pessoas neste mundo acreditam que, após a morte, enfrentarão um sujeito de olhos severos chamado São Pedro, que examinará suas ações na vida e acumulará uma pontuação para a moralidade. Presumivelmente, a regra de pontuação de São Pedro é única e invariante sob mudanças triviais de perspectiva. Infelizmente, os crentes não podem obter uma especificação quantitativa e precisamente computável da regra de pontuação, o que pare bastante injusto.

A religião da Bayesianidade sustenta que seu destino eterno depende dos julgamentos de probabilidade que você fez na vida. Ao contrário de crenças menores, a Bayesianidade pode dar uma especificação quantitativa e precisamente computável de como seu destino eterno é determinado.

Nossa regra de pontuação bayesiana adequada fornece uma maneira de acumular pontuações em todos os experimentos, e a pontuação é invariante, independentemente de como dividimos os "experimen-

tos" ou em que ordem acumulamos os resultados. Somamos os logaritmos das probabilidades. Isso corresponde a multiplicar a probabilidade atribuída ao resultado em cada experimento, para encontrar a probabilidade conjunta de todos os experimentos juntos. Aplicamos o logaritmo para simplificar nossa compreensão intuitiva da pontuação acumulada, para manter nosso controle sobre as minúsculas frações envolvidas e para garantir que maximizamos nossa pontuação esperada declarando nossas probabilidades honestas em vez de colocar todo o nosso dinheiro de jogo na aposta mais provável.

A Bayesianidade afirma que, quando você morre, Pierre-Simon Laplace examina cada evento em sua vida, desde encontrar seus sapatos ao lado da cama pela manhã até encontrar seu local de trabalho em seu local habitual. Cada bilhete de loteria perdido significa que você se importou o suficiente para jogar. Laplace avalia a probabilidade antecipada que você atribuiu a cada evento. Onde você não atribuiu uma probabilidade numérica precisa com antecedência, Laplace examina seu grau de antecipação ou surpresa, extrapola outros resultados possíveis e suas reações extrapoladas e renormaliza suas emoções extrapoladas para uma distribuição de verossimilhança sobre os resultados possíveis. (Daí a frase "superinteligência laplaciana".)

Então Laplace pega cada evento em sua vida e cada probabilidade que você atribuiu a cada evento e multiplica todas as probabilidades juntas. Este é o seu Julgamento Final - a probabilidade que você atribuiu à sua vida.

Aqueles que seguem a Bayesianidade se esforçam por toda a vida para maximizar seu Julgamento Final. Esta é a única virtude da Bayesianidade. O resto é apenas matemática.

Observe: o caminho da Bayesianidade é estrito. Qual probabilidade você deve atribuir a cada manhã à proposição: "O Sol nascerá?" (Descontaremos questiúnculas como dias nublados que a Terra orbita o Sol.) Talvez alguém que não seguisse a Bayesianidade fosse humilde e desse uma probabilidade de 99,9%. Mas nós, que seguimos a Bayesianidade, devemos descartar todas as considerações de modéstia e arrogância, e planejar apenas maximizar nosso Julgamento Final. Como um jogador obsessivo de videogame, nos preocupamos apenas com essa pontuação numérica. Enfrentemos esse problema do nascer do sol 365 vezes por ano, então podemos melhorar consideravelmente nosso Julgamento Final ajustando nossa atribuição de probabilidade.

Do jeito que está, mesmo que o Sol nasça todas as manhãs, a cada ano nosso Julgamento Final diminuirá por um fator de $0{,}999^{365} = 0{,}7$, aproximadamente -0: 52 bits. A cada dois anos, nosso Julgamento Final diminuirá mais do que se nos encontrássemos ignorantes do resultado do lançamento de uma moeda! Intolerável. Se aumentarmos nossa probabilidade diária de nascer do sol para 99,99%, então a cada ano nosso Julgamento Final diminuirá apenas por um fator de 0,964. Melhor. Ainda assim, no caso improvável de vivermos exatamente 70 anos e depois morrermos, nosso Julgamento Final será apenas 7,75% do que poderia ter sido. E se atribuirmos uma probabilidade de 99,999% ao nascer do sol? Então, depois de 70 anos, nosso Julgamento Final será multiplicado por 77,4%.

Por que não atribuir uma probabilidade de 1,0?

Quem segue a Bayesianidade nunca atribuirá uma probabilidade de 1,0 a nada. Atribuir uma probabilidade de 1,0 a algum resultado usa toda a sua massa de probabilidade. Se você atribuir uma probabilidade de 1,0 a algum resultado e a realidade fornecer uma resposta diferente, você deve ter atribuído ao resultado real uma probabilidade de zero. Este é o único pecado mortal da Bayesianidade. Zero vezes qualquer coisa é zero. Quando Laplace multiplicar todas as probabilidades de sua vida, a probabilidade combinada será zero. Seu Julgamento Final será nada, nadica, nadinha, nulo. Não importa o quão racionais sejam seus palpites durante o resto de sua vida, você passará a eternidade ao lado de algum cara que acreditava em discos voadores e obteve todas as suas informações do Weekly World News. Novamente, achamos útil tomar o logaritmo, revelando o inocente "zero" em sua verdadeira forma. Arriscar uma probabilidade de resultado zero é como aceitar uma aposta com um pagamento de infinito negativo.

E se a humanidade decidir desmontar o Sol para obter massa (engenharia estelar) ou desligar o Sol porque está desperdiçando entropia? Bem, você diz, você verá isso chegando, você terá a chance de alterar sua atribuição de probabilidade antes do evento real. E se uma Inteligência Artificial no porão de alguém se auto-aperfeiçoar recursivamente para a superinteligência, desenvolver furtivamente a nanotecnologia e uma

manhã desmontar o Sol? Se na última noite do mundo você atribuir uma probabilidade de 99,999% ao nascer do sol de amanhã, seu Julgamento Final diminuirá por um fator de 100.000. Menos 50 decibéis! Horrível, não é?

Então, qual é a sua melhor estratégia? Bem, suponha que você antecipe 50% que uma superinteligência de IA gerada no porão desmontará o Sol em algum momento nos próximos dez anos, e você calcula que há uma chance igual de isso acontecer em qualquer dia entre agora e então. Em qualquer noite, você anteciparia 99,98% que o Sol nascerá amanhã. Se isso é realmente o que você antecipa, então você não tem motivo para dizer nada, exceto 99,98% como sua probabilidade. Se você se sentir nervoso porque essa antecipação é muito baixa ou muito alta, não deve ser o que você antecipa depois que seu nervosismo for considerado.

Mas a verdade mais profunda da Bayesianidade é esta: você não pode enganar o sistema. Você não pode dar uma resposta humilde, nem uma confiante. Você deve descobrir exatamente o quanto você antecipa que o Sol nascerá amanhã e dizer esse número. Você deve raspar cada fio de cabelo de modéstia ou arrogância e perguntar se espera acabar sendo pontuado no nascer do sol ou na falha em nascer. Não olhe para suas desculpas, mas pergunte quais desculpas você espera precisar. Depois de chegar ao seu grau exato de antecipação, a única maneira de melhorar ainda mais seu Julgamento Final é melhorar a precisão, calibração e discriminação de sua antecipação. Você não pode fazer melhor, exceto adivinhando melhor e antecipando com mais precisão.

Er, bem, exceto que você poderia cometer suicídio quando fizesse cinco anos, impedindo assim que seu Julgamento Final diminuísse ainda mais. Ou se colocarmos um novo pecado na função de utilidade, ordenando contra o suicídio, você poderia fugir do mistério, evitando todas as situações em que pensasse que poderia não saber tudo. Tanto para essa religião.

Idealmente, prevemos o resultado do experimento com antecedência, usando nosso modelo, e então realizamos o experimento para ver se o resultado corresponde com nosso modelo. Infelizmente, nem sempre podemos controlar o fluxo de informações. Às vezes, a Natureza nos lança experiências e, quando pensamos em uma explicação, já vimos os dados que deveríamos explicar. Este foi um dos pecados científicos cometidos pelo evolucionismo do século XIX; Darwin observou a semelhança de muitas espécies e sua adaptação a ambientes locais específicos, antes que a hipótese da seleção natural lhe ocorresse. O evolucionismo do século XIX começou sua vida como uma explicação post facto, não como uma previsão antecipada.

Isso não é um problema apenas de teorias semi-técnicas. Em 1846, a dedução bem-sucedida da existência de Netuno a partir de perturbações gravitacionais na órbita de Urano foi considerada um grande triunfo para a teoria da gravitação de Newton. Por quê? Porque a existência de Netuno foi a primeira observação que confirmou uma previsão antecipada da gravitação newtoniana. Todos os outros fenômenos que Newton explicou, como órbitas e perturbações orbitais e marés, foram observados em grande detalhe antes que Newton os explicasse. Ninguém duvidava seriamente que a teoria de Newton estivesse correta. A teoria de Newton explicava muito precisamente, e substituiu uma coleção de modelos improvisados por uma única lei matemática unificada. Mesmo assim, a previsão antecipada da existência de Netuno, seguida pela observação de Netuno em quase exatamente o local previsto, foi considerada o primeiro grande triunfo da teoria de Newton em prever o que nenhum modelo anterior poderia prever. Um tempo considerável se passou entre a aceitação generalizada da teoria de Newton e a primeira previsão antecipada impressionante da gravitação newtoniana. Quando Newton elaborou sua teoria, os cientistas já haviam observado, em grande detalhe, a maioria dos fenômenos que a gravitação newtoniana previa.

Mas a regra da previsão antecipada é uma moralidade da ciência, não uma lei da teoria da probabilidade. Se você já viu os dados que deve explicar, então a Ciência pode condená-lo ao inferno, mas sua situação não entra em colapso com as leis da teoria da probabilidade. O que acontece é que se torna muito mais difícil para um humano infeliz obedecer às leis da teoria da probabilidade. Quando você está decidindo como classificar uma hipótese segundo a regra de pontuação bayesiana, você precisa descobrir quanta massa de probabilidade essa hipótese atribui ao resultado observado. Se devemos fazer nossas previsões com antecedência, então é mais fácil perceber quando alguém está tentando reivindicar todos os resultados possíveis como uma previsão antecipada, usando muita massa de probabilidade, sendo deliberadamente vago para evitar a falsificação e assim por diante.

Nenhum numerólogo pode prever os números vencedores da loteria da próxima semana, mas eles ficarão felizes em explicar o significado místico dos números vencedores da loteria da semana passada. Digamos que a Mega Ball vencedora foi sete na loteria da semana passada, de 52 resultados possíveis. Obviamente, isso aconteceu porque sete é o número da sorte. Então, a Mega Ball na loteria da próxima semana também será sete? Entendemos que não é certo, é claro, mas se for o número da sorte, você deve atribuir uma probabilidade maior que 1/52… e então pontuaremos seus palpites ao longo de alguns anos, e se sua pontuação for muito baixa, vamos açoitá-lo... o que você disse? Você quer atribuir uma probabilidade de exatamente 1/52? Mas essa é a mesma probabilidade de todos os outros números; o que aconteceu com o sete ser o número da sorte? Não, desculpe, você não pode atribuir uma probabilidade de 90% a sete e também uma probabilidade de 90% a onze. Entendemos que ambos são números da sorte. Sim, entendemos que eles são números de muita sorte. Mas não é assim que funciona.

Mesmo que o ouvinte não conheça o caminho de Bayes e não peça probabilidades formais, ele ficará provavelmente desconfiado se você tentar cobrir muitas bases. Suponha que eles peçam que você preveja a Mega Ball vencedora da próxima semana, e você use numerologia para explicar por que a bola número um se encaixaria muito bem em sua teoria, e por que a bola número dois se encaixaria muito bem em sua teoria, e por que a bola número três se encaixaria muito bem em sua teoria… mesmo o ouvinte mais crédulo pode começar a fazer perguntas quando você chegar a doze. Talvez você possa nos dizer quais números são azarados e definitivamente não ganharão na loteria? Bem, treze é azar, mas não é absolutamente impossível (você se protege, antecipando com antecedência qual desculpa você pode precisar).

Mas se pedirmos que você explique os números da loteria da semana passada, por que, o sete era praticamente inevitável. Aquele sete definitivamente deve contar como um grande sucesso para o modelo de "números da sorte" da loteria. E não poderia ter sido treze; a teoria da sorte descarta isso completamente.

Imagine que você acorda uma manhã e seu braço esquerdo foi substituído por um tentáculo azul. O tentáculo azul obedece aos seus comandos motores - você pode usá-lo para pegar óculos, dirigir um carro, etc. Como você explicaria esse cenário hipotético? Reserve um momento para ponderar sobre este enigma antes de continuar.

(Aviso de spoiler…)

Como eu explicaria o evento de meu braço esquerdo ser substituído por um tentáculo azul? A resposta é que eu não faria. Não acontecerá

Seria fácil produzir uma explicação verbal que "se ajustasse" à hipótese. Existem muitas explicações que podem "se ajustar" em qualquer coisa, incluindo (como um caso especial de "qualquer coisa") meu braço ser substituído por um tentáculo azul. A intervenção divina é uma boa explicação para todos os fins. Ou alienígenas com motivos e capacidades arbitrários. Ou eu poderia estar louco, alucinando, sonhando minha vida em um hospital. Essas explicações "se ajustam" em todos os resultados igualmente bem e igualmente mal, equivalendo a hipóteses de completa ignorância.

O teste para saber se um modelo de realidade "explica" meu braço se transformando em um tentáculo azul é se o modelo concentra massa de probabilidade significativa nesse resultado específico. Por que esse sonho, no hospital? Por que os alienígenas fariam essa coisa específica comigo, em oposição às outros bilhões de coisas que eles poderiam fazer? Por que meu braço se transformaria em um tentáculo naquela manhã, após permanecer um braço em todas as outras manhãs da minha vida? E em todos os casos, devo procurar um argumento convincente o suficiente para fazer essa previsão específica com antecedência, não mera compatibilidade. Uma vez que eu já soubesse o resultado, seria muito mais difícil peneirar as hipóteses para encontrar boas explicações. Independentemente da hipótese que eu tentasse, eu teria dificuldade em não alocar mais massa de probabilidade ao resultado do tentáculo azul de ontem do que se eu extrapolasse cegamente, buscando a previsão mais provável do modelo para amanhã.

Um modelo nem sempre prevê todas as características dos dados. A natureza não tem tendência privilegiada a me apresentar desafios solucionáveis. Talvez uma divindade brinque comigo, e a mente da divindade seja computacionalmente intratável. Se eu jogar uma moeda justa, não há como explicar melhor o resultado, nenhum modelo que faça uma previsão melhor do que a hipótese de entropia máxima. Mas se

eu adivinhar um modelo sem detalhes internos ou um modelo que não faz mais previsões, não só não tenho razão para acreditar nesse palpite, como também não tenho razão para me importar. Ontem à noite, meu braço foi substituído por um tentáculo azul. Por quê? Alienígenas! Então, o que eles farão amanhã? Da mesma forma, se eu atribuir o tentáculo azul a uma alucinação enquanto sonho minha vida em coma, ainda não sei mais sobre o que alucinarei amanhã. Então, por que eu me importo se foram alienígenas ou alucinação?

O que poderia ser uma boa explicação, então, se eu acordasse uma manhã e encontrasse meu braço transformado em um tentáculo azul? Reivindicar uma "boa explicação" para essa experiência hipotética exigiria um argumento tal que, contemplando o argumento hipotético agora, antes que meu braço se transforme em um tentáculo azul, eu fosse dormir preocupado que meu braço realmente se transformasse em um tentáculo.

As pessoas brincam com a plausibilidade, explicando os eventos que esperam nunca encontrar, mas isso necessariamente viola as leis da teoria da probabilidade. Quantas pessoas que pensaram que poderiam "explicar" a experiência hipotética de acordar com o braço substituído por um tentáculo iriam dormir se perguntando se isso realmente poderia acontecer com elas? Se tivessem a coragem de suas convicções, diriam: não espero nunca encontrar essa experiência hipotética e, portanto, não posso explicar, nem tenho motivo para tentar. Essas coisas só acontecem em webcomics, e não preciso preparar explicações, pois na realidade nunca terei a chance de usá-las. Se eu me encontrar nesta situação impossível, deixe-me não perder um jota ou til da minha valiosa perplexidade.

Para um bayesiano, as probabilidades são antecipações, não meras crenças para proclamar dos telhados. Se tenho um modelo que atribui massa de probabilidade a acordar com um tentáculo azul, então estou nervoso sobre acordar com um tentáculo azul. E se o modelo for fantasioso, como uma bruxa lançando um feitiço que me transporta para um webcomic selecionado aleatoriamente? Então, a probabilidade anterior de feitiçaria de webcomic é tão baixa que meu entendimento da realidade não atribui nenhum peso significativo a essa hipótese. A hipótese da feitiçaria, se tomada como um dado adquirido, pode atribuir probabilidade não insignificante a acordar com um tentáculo azul. Mas minha antecipação dessa hipótese é tão baixa que não antecipo nenhuma das previsões dessa hipótese. O fato de eu poder conceber uma hipótese de feitiçaria não deve de forma alguma diminuir minha perplexidade absoluta se eu acordar realmente com um tentáculo, porque a probabilidade real que atribuo à hipótese de feitiçaria é efetivamente zero. Minha hipótese de probabilidade zero não me ajudaria a explicar o acordar com um tentáculo, porque o argumento não é bom o suficiente para me fazer antecipar acordar com um tentáculo.

Nas leis da teoria da probabilidade, as distribuições de verossimilhança são propriedades fixas de uma hipótese. Na arte da racionalidade, explicar é antecipar. Antecipar é explicar. Suponha que eu seja um pesquisador médico e, no curso normal de minha pesquisa, percebo que minha nova teoria inteligente de anatomia parece permitir uma pequena e vaga possibilidade de que meu braço se transforme em um tentáculo azul. "Ha! ha!" Eu digo, "que notável e bobo!" e me sinto um pouco nervoso. Essa seria uma boa explicação para acordar com um tentáculo, se isso acontecesse.

Se uma cadeia de raciocínio não me deixa nervoso, com antecedência, sobre acordar com um tentáculo, então esse raciocínio seria uma explicação ruim se o evento realmente acontecesse, porque a combinação de probabilidade anterior e verossimilhança era muito baixa para me fazer alocar qualquer massa de probabilidade significativa do mundo real para esse resultado.

Se você começar com antecedentes bem calibrados e aplicar o raciocínio bayesiano, terminará com conclusões bem calibradas. Imagine que dois milhões de entidades, espalhadas por diferentes planetas do universo, tenham a oportunidade de encontrar algo tão estranho quanto acordar com um tentáculo (ou - arf! - dez dedos). Um milhão dessas entidades diz "um em mil" para a probabilidade anterior de alguma hipótese X, e cada hipótese X diz "um em cem" para a probabilidade de acordar com um tentáculo. E um milhão dessas entidades diz "um em cem" para a probabilidade anterior de alguma hipótese Y, e cada hipótese Y diz "um em dez" para a probabilidade de acordar com um tentáculo. Se supormos que todas as entidades são bem calibradas, então veremos o universo e encontraremos dez entidades que acabaram com um tentáculo devido a hipóteses de classe de plausibilidade X, e mil entidades que acabaram com tentáculos devido a hipóteses de classe de plausibilidade Y. Então, se você se encontrar com um tentáculo, e se suas probabilidades forem

bem calibradas, então é mais provável que o tentáculo provenha de uma hipótese que você classificaria como provável do que uma hipótese que você classificaria como improvável. (E se suas probabilidades forem mal calibradas, de modo que quando você diz "milhão para um" isso acontece uma vez em vinte? Então você está extremamente confiante e ajustamos suas probabilidades na direção de menos discriminação e maior entropia.)

A hipótese de ser transportado para um webcomic, mesmo que "explique" o cenário de acordar com um tentáculo azul, é uma explicação ruim devido a sua baixa probabilidade anterior. A hipótese do webcomic não contribui para explicar o tentáculo, porque não o faz antecipar acordar com um tentáculo.

Se começarmos com um quatrilhão de mentes sencientes espalhadas pelo universo, muitas entidades encontrarão eventos que são muito prováveis, apenas cerca de um milhão de entidades experimentarão eventos com probabilidades de vida de um bilhão para um (como anteciparíamos, pesquisando com olhos infinitos e calibração perfeita), e nenhuma entidade experimentará o impossível.

Se, de alguma forma, você acordou realmente com um tentáculo, provavelmente seria devido a algo muito mais provável do que "ser transportado para um webcomic", alguma razão perfeitamente normal para acordar com um tentáculo que você simplesmente não viu chegando. Uma razão como o quê? Eu não sei. Nada. Não antecipo acordar com um tentáculo, então não posso dar nenhuma boa explicação para isso. Por que eu deveria me preocupar em criar desculpas que não espero usar? Se eu estivesse preocupado de que um dia pudesse precisar de uma desculpa inteligente para acordar com um tentáculo, a razão pela qual eu estava nervoso com a possibilidade seria minha explicação.

A realidade distribui experiências usando probabilidade, não plausibilidade. Se você descobrir que seu laptop não obedece à Conservação do Momento, então a realidade deve pensar que é uma coisa perfeitamente normal de se fazer com você. Como violar a Conservação do Momento pode ser perfeitamente normal? Antecipo que essa pergunta não tem resposta e nunca precisará ser respondida. Da mesma forma, as pessoas não acordam com tentáculos, então aparentemente não é perfeitamente normal.

Há uma verdade chocante, tão surpreendente e aterrorizante que as pessoas resistem às implicações com todas as suas forças. No entanto, existem alguns solitários com a coragem de aceitar este satori. Aqui está a sabedoria, se você quiser ser sábio:

Desde o início

Nem uma coisa incomum

Jamais aconteceu.

Ai daqueles que desviam os olhos das zebras e sonham com dragões! Se não pudermos aprender a ter alegria no meramente real, nossas vidas serão vazias de fato.

Referências

[1] Edwin T. Jaynes, Probability Theory: The Logic of Science, ed. George Larry Bretthorst (New York: Cambridge University Press, 2003), doi:10.2277/0521592712.

[2] Feynman, Leighton, and Sands, The Feynman Lectures on Physics.

[3] Leitores com conhecimentos de cálculo podem verificar que, no caso mais simples de uma luz que tem apenas duas cores, com p sendo a aposta na primeira cor e f a frequência da primeira cor, o pagamento esperado $f \times (1 - (1 - p)^2) + (1 - f) \times (1 - p^2)$, com p variável e f constante, tem seu máximo global quando definimos p = f.

[4] Não se lembra de como ler P(A|B)? Consulte Uma explicação intuitiva do raciocínio bayesiano.

[5] J. Frank Yates et al., "Probability Judgment Across Cultures," in Gilovich, Griffin, and Kahneman, Heuristics and Biases, 271–291.

[6] Karl R Popper, The Logic of Scientific Discovery (New York: Basic Books, 1959).

[7] Jaynes, Probability Theory.

[8] Imagination Engines, Inc., "The Imagination Engine® or Imagitron™," 2011, www.imagination-engines.com/ie.htm.

[9] Friedrich Spee, Cautio Criminalis; or, A Book on Witch Trials, ed. and trans. Marcus Hellyer, Studies in Early Modern German History (1631; Charlottesville: University of Virginia Press, 2003).

[10] Citado em Dave Robinson and Judy Groves, Philosophy for Beginners, 1st ed. (Cambridge: Icon Books, 1998).

[11] TalkOrigins Foundation, "Frequently Asked Questions about Creationism and Evolution," www.talkorigins.org/origins/faqs-qa.html.

[12] Daniel C. Dennett, Darwin's Dangerous Idea: Evolution and the Meanings of Life (Simon & Schuster, 1995).

[13] Citado em Lyle Zapato, "Lord Kelvin Quotations," 2008, http://zapatopi.net/kelvin/quotes/.

[14] Charles Darwin, On the Origin of Species by Means of Natural Selection; or, The Preservation of Favoured Races in the Struggle for Life, 1st ed. (London: John Murray, 1859), http://darwin- online.org.uk/content/frameset?viewtype=text&itemID=F373&pageseq=1; Charles Darwin, The Descent of Man, and Selection in Relation to Sex, 2nd ed. (London: John Murray, 1874), http://darwin-online.org.uk/content/frameset?itemID=F944&viewtype=text&pageseq=1.

[15] Williams, Adaptation and Natural Selection.

[16] Carl Sagan, The Demon-Haunted World: Science as a Candle in the Dark, 1st ed. (New York: Random House, 1995).

[17] Kevin Brown, Reflections On Relativity (Raleigh, NC: impresso pelo autort, 2011), 405-414, www.mathpages.com/rr/rrtoc.htm.

[18] Ibid.

[19] Stephen Thornton, "Karl Popper," in The Stanford Encyclopedia of Philosophy, Winter 2002, ed. Edward N. Zalta (Stanford University), http://plato.stanford.edu/archives/win2002/entries/ popper/.

[20] John Baez, "The Crackpot Index," 1998, http://math.ucr.edu/home/baez/crackpot.html.