

*LIVRO 2*  
**COMO MUDAR DE IDEIA  
DE VERDADE**

**RACIONALIDADE**

De A a Z

**ELIEZER YUDKOWSKY**

# **RACIONALIDADE DE A a Z**

*COMO MUDAR DE IDEIA DE VERDADE*

*LIVRO 2*

por ELIEZER YUDKOWSKY

Tradução de Mariana Hungria

Brasil, 2024

# Sumário

Racionalidade: uma introdução	6
<b><u>E - Desculpas excessivamente convenientes</u></b>	
46 - O uso adequado da humildade	13
47 - A Terceira Alternativa	16
48 - Loterias: um desperdício de esperança	18
49 - Nova loteria aprimorada	19
50 - Mas ainda há uma chance, certo?	20
51 - A falácia do cinza	22
52 - Autoridade absoluta	25
53 - Como me convencer de que $2 + 2 = 3$	29
54 - Certeza Infinita	31
55 - 0 e 1 não são probabilidades	34
56 - Sua Racionalidade é da Minha Conta	37
<b><u>F - Política e Racionalidade</u></b>	
57 - A política é o assassino da mente	40
58 - Debates sobre políticas não devem parecer unilaterais	41
59 - A balança da Justiça, o caderno da Racionalidade	43
60 - Viés de correspondência	45
61 - Seus inimigos são naturalmente maus?	47
62 - A estupidez ao contrário não é inteligência	49
63 - Argumento bloqueia autoridade	51
64 - Abrace a pergunta	55
65 - A Racionalidade e a língua inglesa	57
<b><u>G - Contra a racionalização</u></b>	
66 - Maldade humana e pensamento confuso	61
67 - Saber sobre vieses pode machucar as pessoas	64
68 - Atualize-se progressivamente	66
69 - Um argumento contra um exército	68
70 - O resultado final	70
71 - Que evidências filtraram evidências?	72
72 - Racionalização	74

73 - Um argumento racional	76
74 - Evitando os verdadeiros pontos fracos de sua crença	78
75 - Parada motivada e continuação motivada	81
76 - Justificativas falsas	83
77 - É essa a sua verdadeira objeção?	85
78 - Verdades Entrelaçadas, Mentiras Contagiosas	87
79 - Sobre mentiras e explosões de Cisne Negro	90
80 - Epistemologia do lado negro	91

## **H - Contra o duplipensar**

81 - Pensamento único	95
82 - Duplipensar (escolhendo ser tendencioso)	97
83 - Não, realmente, eu me autoenganei	100
84 - Crença no autoengano	101
85 - O Paradoxo de Moore	104
86 - Não acredite que você vai se autoenganar	106

## **I - Ver com novos olhos**

87 - Ancoragem e Ajuste	109
88 - Pré-ativação e Contaminação	111
89 - Acreditamos em tudo o que nos dizem?	113
90 - Pensamentos armazenados em <i>cache</i>	115
91 - A Caixa "Fora da Caixa"	117
92 - Visão original	119
93 - A realidade é mais estranha que a história	121
94 - A falácia lógica da generalização a partir de evidências fictícias	122
95 - A virtude da especificidade	126
96 - Como parecer (e ser) profundo	128
97 - Mudamos de ideia com menos frequência do que pensamos	130
98 - Adiar propor soluções	131
99 - A falácia genética	133

## **J - Espirais de autodestruição**

100 - A heurística do afeto	136
101 - Avaliabilidade (e compras de Natal baratas)	138
102 - Escalas ilimitadas, enormes prêmios do júri e futurismo	141
103 - O efeito Halo	143

104 - Viés de super-herói	146
105 - Meros Messias	148
106 - Espirais de autodestruição afetiva	150
107 - Resista à espiral da autodestruição feliz	152
108 - Supercriticidade acrítica	155
109 - Resfriamento evaporativo das crenças de grupos	157
110 - Quando ninguém ousa pedir moderação	159
111 - O experimento da caverna dos ladrões	161
112 - Toda Causa Quer Ser Uma Seita	163
113 - Guardiões da verdade	165
114 - Os guardiões do pool genético	168
115 - Os guardiões de Ayn Rand	169
116 - Dois <i>Koans</i> de Seita	173
117 - Experimento de Conformidade de Asch	175
118 - Sobre Como Expressar Suas Preocupações	178
119 - Dissidência solitária	180
120 - Anti-sectarismo sectário	183

## **K - Deixar ir**

121 - A importância de dizer “Opa!”	189
122 - A oferta absurda	191
123 - Simplesmente perca as esperanças de uma vez	193
124 - O uso adequado da dúvida	194
125 - Você pode encarar a realidade	196
126 - A Meditação sobre a Curiosidade	197
127 - Ninguém pode isentá-lo das leis da racionalidade	200
128 - Deixe uma linha de retirada	202
129 - Crise de fé	205
130 - O Ritual	210

# Racionalidade: uma introdução

por Rob Bensinger



O que devo acreditar?

Acontece que essa pergunta tem uma resposta certa.

Existe uma resposta correta quando lidamos com incertezas, não apenas quando possuímos uma prova conclusiva. Sempre há um nível apropriado de confiança em uma declaração, mesmo que pareça uma “crença pessoal” e não um “fato” confirmado por um especialista.

No entanto, frequentemente falamos como se a existência de incerteza e discordância transformasse as crenças em uma questão meramente de preferência. Dizemos coisas como “essa é apenas a minha opinião” ou “você tem direito à sua opinião”, como se as afirmações científicas e matemáticas existissem em um plano superior e diferente das crenças meramente “privadas” ou “subjetivas”. Mas, como [Robin Hanson escreve](#):<sup>1 [1]</sup>

Você nunca tem direito à sua opinião. Nunca! Você nem mesmo tem direito a dizer “eu não sei”. Você tem direito aos seus desejos e, às vezes, às suas escolhas. Você pode ter a liberdade de escolher suas preferências e, se puder fazê-lo, terá o direito de exercê-la. Mas suas crenças não se tratam de você; elas se tratam do mundo. Suas crenças devem ser a sua melhor estimativa disponível sobre como as coisas são; tudo o mais é uma mentira. [...]

É verdade que em alguns assuntos, os especialistas possuem mecanismos mais robustos para resolver disputas. Em outros tópicos, nossos preconceitos e a complexidade do mundo dificultam alcançar conclusões sólidas. [...]

No entanto, nunca se esqueça de que, em qualquer questão sobre como as coisas são (ou como deveriam ser), e em qualquer situação de informação, sempre existe uma melhor estimativa. Você só tem o direito de se esforçar honestamente para encontrar a melhor estimativa possível; tudo o mais é uma mentira.<sup>1</sup>

Vamos supor que você descubra que uma em cada seis pessoas tem uma queda por você — talvez você receba uma carta de um admirador secreto e esteja certo de que é uma das seis pessoas — mas você não tem ideia de qual delas seja. Seu colega de classe Bob é uma das seis possibilidades, mas você não possui nenhuma evidência especial que o favoreça ou o desfavoreça como sendo o interessado. Nesse caso, a probabilidade de Bob ser o apaixonado é de 1:5.

Como existem seis possibilidades, um palpite resultaria em acertar uma vez para cada cinco vezes que errar, em média. Isso é o que queremos dizer quando falamos em “as chances são de 1:5”. Não é correto

---

1 NT: Texto original em inglês. *You are never entitled to your opinion. Ever! You are not even entitled to “I don’t know.” You are entitled to your desires, and sometimes to your choices. You might own a choice, and if you can choose your preferences, you may have the right to do so. But your beliefs are not about you; beliefs are about the world. Your beliefs should be your best available estimate of the way things are; anything else is a lie. [... ] It is true that some topics give experts stronger mechanisms for resolving disputes. On other topics our biases and the complexity of the world make it harder to draw strong conclusions.[...] But never forget that on any question about the way things are (or should be), and in any information situation, there is always a best estimate. You are only entitled to your best honest effort to find that best estimate; anything else is a lie.*

dizer: “Bem, não faço ideia de quem tem uma queda por mim; talvez seja o Bob, talvez não. Então, vou dizer que as chances são de cinquenta por cento.” Mesmo que você prefira dizer “não sei” ou “talvez” e parar por aí, a resposta ainda é 1:5. [2]

Suponhamos também que você tenha notado que as pessoas piscam para você dez vezes mais quando estão interessadas em você. Se o Bob, por exemplo, piscar para você, isso é uma nova evidência. Nesse caso, seria um erro ser cético em relação ao fato de o Bob ser seu admirador secreto; as chances de 10:1 a favor de “uma pessoa aleatória que pisca para mim está interessada em mim” superam as chances de 1:5 contra “o Bob está interessado em mim”.

Também seria um erro dizer: “Essa evidência é tão forte que posso apostar que é o Bob quem está apaixonado por mim! A partir de agora, presumirei que o Bob está interessado em mim.” O excesso de confiança é tão prejudicial quanto a falta de confiança.

Na verdade, há apenas uma resposta possível para essa pergunta que seja matematicamente consistente. Para atualizar nossas estimativas anteriores de 1:5 com base na razão de probabilidade de 10:1 da evidência, multiplicamos os numeradores juntos e os denominadores juntos, obtendo probabilidades posteriores de 10:5 ou probabilidades de 2:1 a favor de ‘o Bob está interessado em mim’. Com base em nossas suposições e na evidência disponível, supor que o Bob está interessado em você estará correto duas vezes a cada vez que estiver errado. Equivalentemente, a probabilidade de ele estar atraído por você é de 2/3. Qualquer outro nível de confiança seria inconsistente.

Nossa cultura ainda não internalizou as lições da teoria da probabilidade — que a resposta correta para perguntas como “Qual é a probabilidade de o Bob estar interessado em mim?” é tão logicamente limitada quanto a resposta correta para uma pergunta em um teste de álgebra ou em um livro de geologia. Nossos clichês estão em desacordo com a descoberta de que “que crenças devo ter?” Tem uma resposta objetivamente correta, seja sua pergunta “Meu colega de classe está interessado em mim?” ou “Eu tenho uma alma imortal?” Realmente existe uma maneira correta de mudar de opinião. E é uma maneira *precisa*.

### ***Como não mudar de ideia de verdade***

Revisar nossas crenças em algo que se assemelhe remotamente a essa forma idealizada é uma tarefa complicada.

No primeiro volume de *Racionalidade: da IA aos zumbis*, discutimos o valor das crenças “adequadas”. Não há nada intrinsecamente errado em expressar apoio a algo que gostamos, como um grupo ao qual nos identificamos ou uma experiência espiritual que consideramos exaltante. No entanto, quando misturamos esse apoio com crenças factuais, esses aplausos mal compreendidos podem ajudar a proteger uma ideologia inteira da contaminação pelas evidências.

Mesmo crenças que parecem explicar elegantemente nossas observações não estão imunes a esse problema. É muito fácil para nós ver uma frase que soa vagamente científica (ou autoritária) e concluir que ela “explicou” algo, mesmo quando não afeta as probabilidades que implicitamente atribuímos às nossas possíveis experiências futuras.

O pior de tudo é que crenças prosaicas — crenças que *são*, em princípio, passíveis de falsificação, crenças que nossas expectativas — ainda podem se fixar em nossas mentes, reforçadas por uma teia de ilusões e preconceitos.

Em 1951, um jogo de futebol entre Dartmouth e Princeton se tornou extraordinariamente conturbado. Os psicólogos Hastorf e Cantril perguntaram aos alunos de cada universidade quem havia começado a briga. Quase todos concordaram que Princeton não havia começado; no entanto, 86% dos alunos de Princeton acreditavam que Dartmouth havia começado, enquanto apenas 36% dos alunos de Dartmouth culpavam a própria Dartmouth. (A maioria dos alunos de Dartmouth acreditava que “ambos começaram”.)

Não há motivo para pensar que isso foi apenas uma questão de alegria em contraste com crenças reais. Provavelmente, os alunos foram influenciados por suas diferentes crenças e fizeram previsões distintas sobre o comportamento dos jogadores em jogos futuros. No entanto, de alguma forma, as crenças factuais perfeitamente comuns em Dartmouth eram completamente diferentes das crenças factuais perfeitamente

comuns em Princeton.

Podemos atribuir essa diferença apenas às diferentes fontes de informação às quais os alunos de Dartmouth e Princeton tiveram acesso? O viés nas fontes de notícias nas quais os grupos confiam é, por si só, um problema sério.

No entanto, há mais do que isso em ação neste caso. Quando os alunos *assistiram* a um filme do jogo mais tarde e solicitados a contar as infrações que viram, os alunos de Dartmouth afirmaram ter visto, em média, 4,3 infrações da equipe de Dartmouth (e consideraram metade delas como “leves”), enquanto os alunos de Princeton afirmaram ter visto, em média, 9,8 infrações de Dartmouth (e consideraram um terço delas como “leves”).

Não importa se facções rivais concordam ou não sobre questões complicadas de política nacional, ou filosofia moral; alunos com lealdades de grupo diferentes não conseguiram concordar nem mesmo sobre o que estavam *vendo*. [3]

Quando algo que valorizamos é ameaçado, seja nossa visão de mundo, nosso grupo, nossa posição social ou qualquer outra coisa — nossos pensamentos e percepções se unem em sua defesa. [4,5] Alguns psicólogos contemporâneos até sugerem que nossa habilidade de apresentar justificativas explícitas para nossas conclusões evoluiu especificamente para nos ajudar a vencer debates. [6]

Um dos *insights* fundamentais da psicologia do século XX, que inspirou desde os seguidores de Freud até os psicólogos cognitivos atuais, é que o comportamento humano frequentemente é impulsionado por processos inconscientes sofisticados. As narrativas que construímos sobre nossos motivos e razões são muito mais tendenciosas e fabricadas do que imaginamos.

Muitas vezes, não percebemos que estamos criando histórias. Quando acreditamos “perceber diretamente” coisas sobre nós mesmos na introspecção, na realidade, frequentemente estamos nos baseando em modelos causais implícitos e frágeis. [7,8] Ao defender nossas crenças, podemos recorrer a um raciocínio instável que não tem relação com como inicialmente chegamos a essas convicções. [9] Em vez de avaliar nossas explicações com base em sua capacidade preditiva, contamos histórias para atribuir sentido ao que pensamos saber.

Como podemos fazer melhor? Como podemos obter uma compreensão realista do mundo, quando nossas mentes são tão propensas à racionalização? Como podemos alcançar uma visão realista de nossa vida mental, quando nossos próprios pensamentos *sobre* o pensamento são suspeitos?

Como podemos nos tornar menos tendenciosos, quando nossos esforços para eliminar preconceitos podem acabar perpetuando seus próprios vieses?

Qual é o ponto de apoio *mais* estável em que poderíamos nos sustentar?

## ***A matemática da racionalidade***

No início do século XX, os matemáticos alcançaram axiomas simples (como os da teoria dos conjuntos) para a aritmética, proporcionando-lhes um critério mais claro para avaliar a precisão de suas conclusões. Quando um ser humano ou uma calculadora produz “ $2 + 2 = 4$ ”, agora podemos ir além do que simplesmente dizer “isso parece intuitivamente correto”. Podemos explicar *por que* está correto e provar que sua validade está sistematicamente relacionada à validade do restante da aritmética.

A matemática e a lógica nos permitem modelar comportamentos de sistemas físicos muito mais complexos do que uma calculadora de bolso. Também podemos formalizar o raciocínio racional de forma geral, utilizando a teoria da probabilidade para identificar características compartilhadas por todas as formas bem-sucedidas de inferência. Além disso, podemos até formalizar o comportamento racional de maneira abrangente, recorrendo à teoria da decisão.

A teoria da probabilidade define como idealmente deveríamos raciocinar diante da incerteza, se tivéssemos tempo, capacidade computacional e autodisciplina ilimitados. Com base em algum conhecimento



prévio e uma nova evidência, a teoria da probabilidade determina de forma única o melhor conjunto de novas crenças que poderíamos adotar (crenças posteriores). Da mesma forma, a teoria da decisão define qual ação devemos tomar com base em nossas crenças. Para qualquer conjunto coerente de crenças e preferências que tenhamos em relação a uma situação, existe uma resposta teórica que nos indica como devemos agir para satisfazer nossas preferências.

Os seres humanos não são raciocinadores nem tomadores de decisão perfeitos, assim como não somos calculadoras perfeitas. Nossos cérebros são adaptações construídas pela seleção natural. Mesmo em nosso melhor desempenho, não calculamos a resposta correta para “o que devo pensar?” e “o que devo fazer?”. Nos falta tempo e poder de processamento, e a evolução não possui a experiência em engenharia e previsão necessária para resolver todos os nossos problemas.

Um raciocinador livre de erros, altamente eficiente no mundo real, ainda precisaria recorrer a heurísticas e aproximações. Os algoritmos computacionalmente otimizados para mudança de crenças não alcançam a consistência total da teoria da probabilidade.

No entanto, mesmo sabendo que não podemos ser totalmente consistentes, certamente podemos melhorar. Saber que existe um padrão ideal com o qual podemos nos comparar — o que os pesquisadores chamam de “racionalidade bayesiana” — pode nos orientar à medida que aprimoramos nossos pensamentos e ações. Embora nunca sejamos bayesianos perfeitos, a matemática da racionalidade pode nos ajudar a entender *por que* uma resposta específica está correta e a identificar precisamente onde erramos.

Imagine tentar aprender matemática apenas por meio de memorização mecânica. Você pode ouvir que “ $10 + 3 = 13$ ”, “ $31 + 108 = 139$ ” e assim por diante, mas isso não será muito útil a menos que você compreenda o padrão subjacente aos números. Pode ser muito mais desafiador buscar métodos para aprimorar sua racionalidade quando você não possui uma estrutura geral para avaliar o sucesso de um método. O propósito deste livro é auxiliar as pessoas a construir para si mesmas tais estruturas.

## ***Racionalidade aplicada***

Em uma postagem de [blog](#) que discute as diferenças entre os entusiastas da racionalidade “racionalistas” e os anti-empiristas “racionalistas”, Scott Alexander observou [\[10\]](#):

“É claro que é útil ter o máximo de evidências possível, assim como é útil ter a maior quantidade possível de dinheiro. No entanto, é igualmente óbvio que é útil saber como utilizar uma quantidade limitada de evidências com sabedoria, da mesma forma que é útil saber como utilizar uma quantidade limitada de dinheiro com sabedoria.”<sup>2</sup>

As técnicas de racionalidade nos ajudam a extrair o máximo de valor das evidências que possuímos nos casos em que estas são inconclusivas ou quando nossos preconceitos e apegos distorcem nossa interpretação. Isso se aplica tanto à nossa vida pessoal, conforme exemplificado na história de João, quanto às divergências entre facções políticas (e entre torcedores de futebol). Também se aplica a quebra-cabeças tecnológicos e filosóficos, como os debates sobre transumanismo, que defende a ideia de remodelar radicalmente a condição humana por meio da tecnologia. Reconhecendo que as mesmas regras matemáticas se aplicam a cada um desses domínios e, que, os mesmos vieses cognitivos prevalecem em muitos casos, *Como Realmente Mudar de Ideia* é embasado em uma ampla gama de exemplos de problemas.

A primeira sequência de ensaios em “Como mudar de ideia de verdade”, intitulada “[Desculpas excessivamente convenientes](#)”, concentra-se em questões tão definidas probabilisticamente quanto possível.

---

2 NT: Texto original em inglês. *Obviously it's useful to have as much evidence as possible, in the same way it's useful to have as much money as possible. But equally obviously it's useful to be able to use a limited amount of evidence wisely, in the same way it's useful to be able to use a limited amount of money wisely.*

Embora calcular a resposta ótima de Bayes seja frequentemente inviável, erros como o viés de confirmação podem se enraizar mesmo quando a evidência disponível é esmagadora e temos muito tempo para pensar nas coisas.

A partir desse ponto, avançamos para assuntos mais complexos com uma sequência sobre “[Política e Racionalidade](#)”. A política nacional dominante, como debatida pelos especialistas da televisão, é conhecida por suas discussões raivosas e improdutivas. Diante disso, há algo surpreendente. Por que levamos as divergências políticas para o lado pessoal, mesmo quando a política nacional está tão distante de nós em termos de espaço e tempo? Além disso, por que não nos tornamos *mais* cuidadosos e rigorosos com as evidências quando lidamos com questões que consideramos importantes?

O jogo entre Dartmouth e Princeton oferece uma resposta. Grande parte do nosso processo de raciocínio é, na verdade, racionalização — uma narrativa que torna nossas crenças atuais mais coerentes e justificadas, sem necessariamente melhorar sua precisão. O ensaio “[Contra a racionalização](#)” aborda esse problema, seguido por “[Contra o duplipensar](#)” (sobre autoengano) e “[Ver com novos olhos](#)” (sobre o desafio de reconhecer evidências que não se encaixam em nossas expectativas e suposições).

Elevar o nível de racionalidade significa encontrar muitas ideias novas e poderosas. Em muitos casos, isso também implica fazer amizades com pessoas com as quais você pode trocar ideias e encontrar comunidades que o incentivem a se aprimorar. O capítulo “[Espiras de autodestruição](#)” discute alguns perigos importantes que podem afetar grupos unidos em torno de interesses comuns e ideias brilhantes incríveis, os quais precisarão ser superados se quisermos colher todos os benefícios das comunidades racionalistas. “*Como mudar de ideia de verdade*” conclui com uma sequência chamada “[Deixar ir](#)”.

Nosso estado natural *não* é mudar de ideia como um bayesiano faria. Convencer os alunos de Dartmouth e Princeton *a enxergarem a realidade como ela é* não será tão fácil quanto recitar os axiomas da teoria da probabilidade para eles. Como Luke Muehlhauser escreve em [The power of agency](#) (O poder da autonomia) [11]:

“Você não é um homúnculo bayesiano cujo raciocínio é “corrompido” por vieses cognitivos. Você é, na verdade, feito de vieses cognitivos.”

O viés de confirmação, o viés de *status quo*, o viés de correspondência e outros não são adicionados ao nosso raciocínio; eles são a própria substância do nosso pensamento.”

Isso não significa que eliminar os vieses seja impossível. Assim como não somos calculadoras perfeitas por trás de todos os nossos erros aritméticos. Muitas das nossas limitações matemáticas decorrem de aspectos profundos do funcionamento do cérebro humano. No entanto, podemos aprimorar nossas habilidades matemáticas; podemos aprender quando confiar e desconfiar de nossas intuições matemáticas, compartilhar nosso conhecimento e nos ajudar mutuamente; podemos moldar nossos ambientes para tornar as coisas mais fáceis para nós e desenvolver ferramentas para aliviar parte do trabalho.

Nossos preconceitos são uma parte intrínseca de nós. No entanto, também existe uma sombra de bayesianismo em nós, um mecanismo falho que pode realmente nos aproximar da verdade. Não somos homúnculos, mas ainda assim há alguma verdade. Talvez seja o suficiente para começar.

---

3 NT: Texto original em inglês. *You are not a Bayesian homunculus whose reasoning is “corrupted” by cognitive biases. You just are cognitive biases.*

## Referências

- [1] Robin Hanson, “You Are Never Entitled to Your Opinion,” *Overcoming Bias* (blog) (2006), [http://www.overcomingbias.com/2006/12/you\\_are\\_never\\_e.html](http://www.overcomingbias.com/2006/12/you_are_never_e.html).
- [2] Isso decorre da suposição de que existem seis possibilidades e você não tem razão para favorecer uma delas em detrimento de qualquer outra. Também estamos supondo, de forma irreal, que você pode realmente ter certeza de que o admirador é uma dessas seis pessoas e que você não está negligenciando outras possibilidades. (E se mais de uma das seis pessoas tiver uma queda por você?)
- [3] Albert Hastorf and Hadley Cantril, “They Saw a Game: A Case Study,” *Journal of Abnormal and Social Psychology* 49 (1954): 129–134, <http://www2.psych.ubc.ca/~schaller/Psyc590Readings/Hastorf1954.pdf>.
- [4] Pronin, “How We See Ourselves and How We See Others.”
- [5] Robert P. Vallone, Lee Ross, and Mark R. Lepper, “The Hostile Media Phenomenon: Biased Perception and Perceptions of Media Bias in Coverage of the Beirut Massacre,” *Journal of Personality and Social Psychology* 49 (1985): 577–585, <http://ssc.wisc.edu/~jpiliavi/965/hwang.pdf>.
- [6] Hugo Mercier and Dan Sperber, “Why Do Humans Reason? Arguments for an Argumentative Theory,” *Behavioral and Brain Sciences* 34 (2011): 57–74, <https://hal.archives-ouvertes.fr/file/index/docid/904097/filename/MercierSperberWhydohumansreason.pdf>.
- [7] Richard E. Nisbett and Timothy D. Wilson, “Telling More than We Can Know: Verbal Reports on Mental Processes,” *Psychological Review* 84 (1977): 231–259, <http://people.virginia.edu/~tdw/nisbett&wilson.pdf>.
- [8] Eric Schwitzgebel, *Perplexities of Consciousness* (MIT Press, 2011).
- [9] Jonathan Haidt, “The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment,” *Psychological Review* 108, no. 4 (2001): 814–834, doi:10.1037/0033-295X.108.4.814.
- [10] Scott Alexander, “Why I Am Not Rene Descartes,” *Slate Star Codex* (blog) (2014), <http://slatestarcodex.com/2014/11/27/why-i-am-not-rene-descartes/>.
- [11] Luke Muehlhauser, “The Power of Agency,” *Less Wrong* (blog) (2011), [http://lesswrong.com/lw/5i8/the\\_power\\_of\\_agency/](http://lesswrong.com/lw/5i8/the_power_of_agency/).



## **E - Desculpas excessivamente convenientes**



## 46 - O uso adequado da humildade



É amplamente reconhecido que a boa ciência requer algum tipo de humildade.

Considere o criacionista que diz: “Mas quem pode realmente saber se a evolução é correta? É apenas uma teoria. Você deveria ser mais humilde e ter a mente aberta.” Isso é humildade? O criacionista pratica uma falta de confiança muito seletiva, recusando-se a integrar pesos maciços de evidências em favor de uma conclusão que eles acham desconfortável. Eu diria que, quer você chame isso de “humildade” ou não, é um passo equivocado na dança.

E o engenheiro que humildemente projeta mecanismos à prova de falhas no maquinário, mesmo estando certo de que o maquinário não falhará? Isso me parece um bom tipo de humildade. Historicamente, não é inédito para um engenheiro ter certeza de que uma nova máquina não falhará, e ela falha de qualquer maneira.

E o aluno que humildemente verifica as respostas em seu teste de matemática? Mais uma vez, eu classificaria isso como boa humildade.

E que tal um aluno que diz: “Bem, não importa quantas vezes eu verifique, nunca posso ter *certeza* de que minhas respostas no teste estão corretas” e, portanto, não verifica nem uma vez? Mesmo que essa escolha decorra de uma emoção semelhante à emoção sentida pelo aluno anterior, ela é menos sábia.

Você sugere estudar mais e o aluno responde: “Não, não daria certo pra mim; não sou tão esperto quanto você; não, alguém tão humilde como eu, não pode esperar nada melhor.” Isso é modéstia social, não humildade. Tem a ver com a regulamentação do status na tribo e não com o processo científico. Se você pedir a alguém para “ser mais humilde”, por padrão, eles associarão as palavras à modéstia social, o qual é um conceito intuitivo, cotidiano e ancestralmente relevante. A humildade científica é uma invenção mais recente e rara, e não é inerentemente social. A humildade científica é algo que você praticaria mesmo se estivesse sozinho em um traje espacial, a anos-luz da Terra, sem ninguém olhando. Ou mesmo se você recebesse uma garantia absoluta de que ninguém jamais o criticaria novamente, não importa o que você dissesse ou pensasse de si mesmo. Você ainda verificaria seus cálculos se fosse sábio.

O aluno diz: “Mas já vi outros alunos verificarem suas respostas e ainda assim eles estavam errados. E se, pelo problema da indução,  $2 + 2 = 5$  desta vez? Não importa o que eu faça, não estarei seguro de mim mesmo.” Parece muito profundo e muito modesto. Mas não é por acaso que o aluno quer entregar a prova rapidinho e ir para casa jogar videogame.

O fim de uma era na física nem sempre se anuncia com trovões e trombetas; na maioria das vezes, começa com o que parece uma pequena falha... Mas como os físicos têm essa ideia arrogante de que seus modelos devem funcionar o tempo *todo*, não apenas a *maioria* do tempo, eles acompanham pequenas falhas. Normalmente, a pequena falha desaparece sob uma inspeção mais detalhada. Raramente, a falha aumenta a ponto de destruir toda a teoria. Por isso está escrito: “Se você não buscar a perfeição, vai parar antes de dar os primeiros passos”.

Mas pense na audácia social de tentar estar *sempre* certo! Suspeito seriamente que, se a ciência afirmasse que a teoria da evolução é verdadeira na maioria das vezes, mas não sempre — ou se a ciência admitisse que talvez, em alguns dias, a Terra *seja* plana, mas quem realmente sabe — os cientistas teriam uma reputação social melhor. A ciência seria vista como menos conflituosa, porque não teríamos que discutir com

peças que afirmam que a Terra é plana — haveria espaço para concessões. Quando você discute muito, as pessoas o consideram agressivo. Se você repetidamente se recusar a fazer concessões, é ainda pior.

Considere isso como uma questão de status tribal: os cientistas certamente ganharam algum status extra em troca de ferramentas socialmente úteis como remédios e telefones celulares. Mas esse status social não justifica sua insistência em que *apenas* ideias científicas sobre a evolução sejam ensinadas nas escolas públicas. Afinal, os padres também têm alto status social. Os cientistas estão se excedendo — ganharam um pouco de status e agora pensam serem os líderes da tribo! Deveriam ser mais humildes e ceder um pouco.

Muitas pessoas parecem ter uma visão um tanto confusa da “humildade racionalista”. É perigoso ter um princípio prescritivo que você compreende apenas vagamente; sua imagem mental pode ter tantas variações que pode ser adaptada para justificar quase qualquer ação. Onde as pessoas possuem modelos mentais vagos que podem ser usados para argumentar qualquer coisa, elas geralmente acabam acreditando em tudo o que desejavam acreditar. Isso é tão conveniente que as pessoas muitas vezes relutam em prescindir da imprecisão. Mas o propósito de nossa ética é nos mover, não sermos movidos por ela.

“Humildade” é uma virtude frequentemente mal compreendida. Isso não significa que devemos descartar o conceito de humildade, mas devemos ter cuidado ao usá-lo. Pode ser útil observar as ações recomendadas por uma abordagem “humilde” e perguntar: “Essa maneira de agir me torna mais forte ou mais fraco?” Se considerarmos o problema da indução aplicado a uma ponte que precisa permanecer de pé, pode parecer razoável concluir que nada é certo, independentemente das precauções tomadas. No entanto, ao considerarmos a diferença no mundo real entre adicionar alguns cabos extras e encolher os ombros, fica claro o que torna a ponte mais forte.

A grande maioria dos apelos que presencio à “humildade do racionalista” são desculpas para encolher os ombros. Aquele que compra um bilhete de loteria, dizendo: “Mas você não pode *saber* que vou perder.” Aquele que não acredita na evolução, dizendo: “Mas você não pode me provar que é verdade.” Aquele que se recusa a enfrentar um problema aparentemente difícil, dizendo: “Provavelmente é muito difícil de resolver.” O problema é o [ceticismo motivado](#), também conhecido como viés de desconfirmação — examinando mais fortemente as afirmações nas quais não queremos acreditar. A humildade, em sua forma mais mal compreendida, é uma desculpa totalmente geral para não acreditar em algo; afinal, você não pode ter certeza. Cuidado com desculpas totalmente gerais!

Outro problema é a humildade ser muito fácil de professar. Dennett, em [Breaking the spell: religion as a natural phenomenon](#) (Quebrando o Feitiço: A Religião como Fenômeno Natural), aponta que, embora muitas afirmações religiosas sejam muito difíceis de acreditar, é fácil para as pessoas acreditarem que *devem* acreditar nelas. Dennett chama isso de “crença na crença”. O que realmente significa supor, realmente acreditar, que três é igual a um? É muito mais fácil acreditar que você *deve*, de alguma forma, acreditar que três é igual a um, e dar essa resposta nos momentos apropriados na igreja. Dennett sugere que grande parte da “crença religiosa” deve ser estudada como “profissão religiosa” — o que as pessoas pensam que devem acreditar e o que sabem que devem dizer.

É muito fácil enfrentar todos os contra-argumentos dizendo: «Bem, é claro que posso estar errado.» Então, tendo obedientemente se curvado em direção à Modéstia, tendo feito a devida reverência, você pode seguir seu caminho sem mudar nada.

A tentação é sempre reivindicar o máximo de pontos com o mínimo de esforço. A tentação é integrar cuidadosamente todas as informações recebidas de uma forma que nos permita mudar nossas crenças e, acima de tudo, nossas ações, o menos possível. John Kenneth Galbraith disse [\[1\]](#): “Diante da escolha de mudar de ideia e provar que não há necessidade de fazê-lo, quase todo mundo fica ocupado com a prova.”<sup>4</sup> E quanto maior a inconveniência de mudar de ideia, mais esforço as pessoas farão para gastar com a prova.

Mas, sabe, se você vai fazer o mesmo de qualquer maneira, não faz sentido ir tão longe para racionalizá-la. Frequentemente, tenho testemunhado pessoas encontrando novas informações, aparentemente aceitando-as e depois explicando cuidadosamente por que farão o mesmo que planejaram fazer anteriormente,

---

4 NT: Texto original em inglês. “Faced with the choice of changing one’s mind and proving that there is no need to do so, almost everyone gets busy on the proof.”

mas com uma justificativa diferente. O objetivo do pensamento é moldar nossos planos; se você vai manter os mesmos planos de qualquer maneira, por que se dar ao trabalho de justificar isso? Quando você encontra novas informações, a parte difícil é atualizar, reagir, em vez de simplesmente deixar as informações desaparecerem em um buraco negro. E a humildade, mal compreendida, cria um maravilhoso buraco negro — tudo o que você precisa fazer é admitir que pode estar errado. Portanto, está escrito: “Ser humilde é tomar ações específicas em antecipação aos próprios erros. Confessar sua falibilidade e depois não fazer nada a respeito não é humildade; é vangloriar-se de sua modéstia”.

## **Referências**

[1] John Kenneth Galbraith, *Economics, Peace and Laughter* (Plume, 1981), 50.

## 47 - A Terceira Alternativa



Acreditar no Papai Noel proporciona às crianças uma sensação de encantamento e as incentiva a se comportarem bem na esperança de receber presentes. Se essa crença for destruída pela verdade, argumenta-se que as crianças perderão o senso de admiração e deixarão de se comportar adequadamente. Portanto, embora a crença no Papai Noel seja falsa, ela é considerada uma “Nobre Mentira” cujo benefício líquido deve ser preservado por razões utilitárias.

Classicamente, esse argumento é um [falso dilema](#), também conhecido como falácia do terceiro excluído ou [falácia do pacote](#) de acordo. Mesmo que aceitemos as premissas factuais e morais subjacentes ao argumento acima, ele permanece inválido. Supondo que a política do Papai Noel (incentivar as crianças a acreditarem nele) seja melhor do que a política nula (não fazer nada), isso não significa que seja a melhor de todas as alternativas possíveis.

Outras abordagens também podem proporcionar às crianças um senso de admiração, como levá-las para assistir ao lançamento de um foguete espacial ou fornecer-lhes romances de ficção científica. Da mesma forma, oferecer recompensas a crianças por bom comportamento as encoraja a se comportarem bem apenas quando os adultos estão observando, enquanto elogios sem recompensas materiais tendem a promover um bom comportamento incondicional.

As “Mentiras Nobres” geralmente são falácias do pacote acordo, e a resposta a uma falácia de pacote de acordo é que, se realmente precisamos do suposto benefício, podemos construir uma Terceira Alternativa para obtê-lo.

Como podemos encontrar essas Terceiras Alternativas? O primeiro passo é decidir procurá-las, e o último é a decisão de aceitá-las. Embora pareça óbvio, a maioria das pessoas falha nessas duas etapas, não no processo de busca em si. De onde surgem os falsos dilemas? Alguns emergem honestamente, pois alternativas superiores podem ser cognitivamente difíceis de visualizar. No entanto, existe uma tendência de criar falsos dilemas para justificar políticas questionáveis, apontando para um suposto benefício em relação à inação. Nesses casos, quem justifica não deseja uma Terceira Alternativa; encontrá-la destruiria a justificativa. A última coisa que um defensor do Papai Noel quer ouvir é que elogios funcionam melhor do que recompensas materiais, ou que naves espaciais podem ser tão inspiradoras quanto renas voadoras.

O ótimo é inimigo do bom. Se o objetivo é realmente ajudar as pessoas, uma alternativa superior deve ser celebrada — assim que encontrarmos uma estratégia melhor, poderemos auxiliar as pessoas com mais eficácia. No entanto, se o objetivo é justificar uma determinada estratégia alegando que ela ajuda as pessoas, uma Terceira Alternativa se torna um [argumento adversário](#), um concorrente.

A psicologia cognitiva moderna enxerga a tomada de decisão como uma busca por alternativas. Na vida real, não basta apenas comparar opções; é necessário gerá-las primeiro. Em muitos problemas, o número de alternativas é vasto e, portanto, é necessário estabelecer um critério de parada para a busca. Quando se deseja comprar uma casa, não é possível avaliar todas as casas da cidade; em algum momento é preciso parar de procurar e decidir.

Mas e quando nossas motivações conscientes para a busca — os critérios que podemos reconhecer em nós mesmos — não estão alinhados com as influências subconscientes? O que acontece quando estamos em uma busca aparentemente altruísta, procurando uma política benéfica, e encontramos uma estratégia que beneficia os outros, mas nos prejudica? Bem, não paramos de procurar *ali*; continuamos buscando. Dize-



mos a nós mesmos que estamos procurando uma estratégia que traga um benefício altruísta maior, é claro. No entanto, se encontrarmos uma política que tenha algum benefício defensável e também seja pessoalmente conveniente, tendemos a parar a busca imediatamente! Na verdade, resistiremos provavelmente a qualquer sugestão de retomar a busca, alegando falta de tempo, talvez. (No entanto, de alguma forma, sempre encontramos recursos cognitivos para justificar nossa política atual.)

Devemos estar atentos quando nos pegarmos argumentando que uma política é defensável em vez de ideal, ou que tem algum benefício em comparação com a inação, em vez do melhor benefício possível.

Falsos dilemas são frequentemente apresentados para justificar políticas antiéticas que, por alguma estranha coincidência, são muito convenientes. Mentir, por exemplo, costuma ser muito mais conveniente do que dizer a verdade; e manter nossas crenças iniciais é mais cômodo do que atualizá-las. É por isso que os argumentos em favor das Mentiras Nobres são populares; eles servem como uma defesa das crenças pré-existentes — não encontramos Mentiras Nobres que calculam uma nova Mentira Nobre ideal; elas simplesmente mantêm qualquer mentira com a qual começaram. É melhor encerrar essa busca rapidamente!

Para fazer melhor, pergunte-se diretamente: *se eu descobrisse haver uma alternativa superior à minha política atual, ficaria genuinamente feliz no fundo do meu coração ou sentiria um leve lampejo de relutância antes de desistir?* Se as respostas forem «não» e «sim», tenha cuidado, pois pode ser que você não tenha realmente buscado uma Terceira Alternativa.

Isso nos leva a outra pergunta importante: você já dedicou cinco minutos com os olhos fechados, pensando em opções criativas e inusitadas, tentando encontrar uma alternativa melhor? É importante que sejam realmente cinco minutos no relógio, porque, caso contrário, você pode apenas piscar os olhos, fechá-los novamente e dizer: “Bem, sim, eu procurei alternativas, mas não encontrei nenhuma”. Piscar os olhos é um [truque conveniente](#) para se livrar de suas responsabilidades. Recomendo o uso de um relógio físico real.

E quanto a essas opções criativas e inusitadas — você teve o cuidado de não descartar nenhuma boa opção? Houve algum esforço secreto no canto da sua mente para garantir que todas as opções consideradas fossem obviamente ruins?

É surpreendente quantos defensores de Mentiras Nobres e seus seguidores estão prontos para abraçar violações éticas — mesmo que expressem pesar e angústia de consciência — quando nem sequer dedicaram cinco minutos no relógio para buscar uma alternativa. Existem certas buscas mentais que secretamente desejamos que falhem; e quando a perspectiva de sucesso é desconfortável, as pessoas usam a primeira desculpa possível para desistir.

## 48 - Loterias: um desperdício de esperança



A crítica tradicional à loteria é que seus jogadores são justamente aqueles com menos recursos para perder. Argumenta-se que a loteria desperdiça dinheiro, drenando a riqueza de quem mais precisa dela. Alguns defensores da loteria, e até mesmo alguns comentaristas especializados em superação de vieses cognitivos, tentaram justificar a compra de bilhetes como uma aquisição racional de fantasia — pagar um real por um dia de expectativa agradável, imaginando-se milionário.

No entanto, é crucial refletir sobre o que isso realmente implica. Significa ocupar sua valiosa mente com uma fantasia cuja probabilidade real de se concretizar é praticamente nula — uma ínfima chance sobre a qual você, por si só, não pode fazer nada para tornar realidade. O futuro será decidido pelas bolas sorteadas na loteria. Essa fantasia alimenta a ideia de riqueza [sem esforço](#) — sem consciência, aprendizado, carisma ou mesmo paciência.

Isso torna a loteria um tipo diferente de sumidouro: um dreno emocional de energia. Ela incentiva as pessoas a investirem seus sonhos e esperanças de um futuro melhor em uma probabilidade minúscula. Se não fosse pela loteria, talvez elas sonhassem em frequentar um curso técnico, abrir seu próprio negócio ou conquistar uma promoção no trabalho — coisas que realmente poderiam fazer, esperanças que as motivariam a se fortalecerem. Em seu estado de devaneio, em sua vigésima representação da fantasia agradável, seus cérebros poderiam vislumbrar uma maneira real de realizar isso. Afinal, para que *servem* os sonhos e o cérebro?

No entanto, como essa perspectiva limitada pela realidade pode competir com a ilusão artificialmente adocicada de riqueza instantânea — não após conduzir uma startup até sua oferta pública inicial, mas sim na próxima terça-feira?

Francamente, por que não podemos simplesmente dizer que comprar bilhetes de loteria é uma atitude tola? Os seres humanos às vezes agem de forma insensata — isso não deveria ser uma suposição tão surpreendente. Não é surpreendente que o cérebro humano não realize cálculos complexos e não consiga diminuir o valor emocional de uma antecipação agradável por um fator de 0,00000001 sem abandonar completamente a linha de raciocínio. Não é surpreendente que muitas pessoas não compreendam que um cálculo numérico da utilidade esperada deveria anular ou substituir seus instintos financeiros imprecisos e, em vez disso, tratem o cálculo apenas como um *argumento* que pode ser contraposto a suas antecipações prazerosas — um argumento emocionalmente fraco, já que é apenas uma mera anotação no papel, em vez de uma visão de riquezas fabulosas.

Isso parece suficiente para explicar a popularidade das loterias. Por que tantos defensores sentem-se compelidos a apoiar essa forma clássica de [autodestruição](#)?

Superar o viés requer um processo que envolve: (1) perceber o viés, (2) analisá-lo em detalhes, (3) decidir que o viés é prejudicial, (4) encontrar uma solução alternativa e (5) implementá-la. É lamentável quantas pessoas concluem as etapas 1 e 2, mas ficam paralisadas na etapa 3, que deveria ser a mais fácil das cinco.

Os vieses são como limões, não limonada, e não devemos tentar transformá-los em limonada — devemos simplesmente deixar esses limões *apodrecerem*.

## 49 - Nova loteria aprimorada



Algumas pessoas [argumentam](#) que a loteria não é um [desperdício de esperança](#), mas um serviço que permite comprar fantasias — “[sonhar acordado](#) em se tornar milionário por um valor muito menor do que ir ao cinema”. Um [comentarista](#) escreveu: “Há uma grande diferença entre chance zero de ficar rico e chance épsilon. Comprar um bilhete preenche essa lacuna.”

Na verdade, entre chance zero e chance épsilon de ficar rico, a diferença é de ordem épsilon. Consideremos épsilon igual a um sobre o Googleplex.

Se a loteria vende esperança épsilon, isso sugere uma Nova Loteria Aprimorada. Ela paga, em média, a cada cinco anos, em momento aleatório. Você compra uma vez, por um dólar, e obtém anos de chance épsilon. A qualquer momento, o telefone pode tocar informando que você virou milionário!

Imagine como seria melhor que uma loteria convencional, com sorteios em horários fixos. Seu chefe exige que você refaça um trabalho chato? Em vez de começar, você poderia olhar para o telefone esperando a ligação — haveria chance épsilon de ganhar naquele exato momento! E se não for agora, pode ser no próximo minuto!

Pense nas fantasias que essa loteria permitiria. Você poderia encher o carrinho de compras com itens caros — se o telefone não tocar, é só devolver tudo, certo?

A Nova Loteria poderia exibir uma distribuição de probabilidade flutuante sobre as chances de vitória, com expectativa seguindo a distribuição de Poisson. Que emoção! “Nossa, agora a chance de ganhar é dez vezes maior!” “O número 42 que escolhi tem o dobro das chances!” Isso poderia aparecer nos celulares das pessoas. Muito mais empolgante que tarefas chatas do dia-a-dia! Um sonho tão irresistível que competiria até com a esperança de sair mais cedo do trabalho. As pessoas ficariam grudadas na tela *o dia todo!*

Oferecer devaneios tentadores é um serviço valioso, certo? As pessoas pagam por isso, então deve ter valor. A alternativa seria pensar que os consumidores cometem erros, mas sabemos que isso é impossível.

Porém, os governos, com seu monopólio cruel das loterias, não oferecem esse serviço óbvio. Por quê? Porque querem explorar as pessoas, fazendo-as gastar dinheiro toda semana. Querem que gastem centenas de dólares pela emoção de sentirem que a chance de ganhar é cem vezes maior, em vez de só olharem a tela esperando a probabilidade aumentar.

Portanto, se você acredita que loteria é um serviço, é evidente que se trata de um serviço caríssimo — cobrado dos mais pobres. É seu dever como cidadão exigir a Nova Loteria Aprimorada.

## 50 - Mas ainda há uma chance, certo?



Há alguns anos, eu estava em uma conversa com alguém quando ele casualmente expressou sua descrença na teoria da evolução. Respondi prontamente: “Não estamos mais no século XIX. Na época em que Darwin apresentou a teoria da evolução, talvez fosse aceitável ter dúvidas. No entanto, estamos no século XXI. Agora podemos decifrar os genes. Por meio desse conhecimento, descobrimos que os seres humanos e os chimpanzés compartilham 98% de seu DNA. Temos a comprovação de que somos parentes próximos dos chimpanzés. Essa questão está resolvida.”

Ele disse: “Talvez o DNA seja apenas semelhante por coincidência”.

Eu disse: “As chances de isso acontecer são algo como 2 elevado à potência de 750 milhões para um”.

E então ele disse: “Mas ainda há uma chance, certo?”

Agora, existem várias razões pelas quais meu eu do passado não pode reivindicar uma vitória moral incontestável nessa conversa. Uma delas é que não consigo me lembrar de onde tirei aquele número,  $2^{750.000.000}$ , embora seja provavelmente a ordem correta de magnitude. Outra razão é que meu eu do passado não considerou o conceito de confiança calibrada. Ao longo da história humana, sempre que alguém calculou chances de  $2^{750.000.000}:1$  contra algo, certamente cometeu erros com mais frequência do que uma vez em  $2^{750.000.000}$ . Por exemplo, a estimativa de genes compartilhados foi revisada para 95%, não 98% — e isso pode se aplicar apenas aos 30.000 genes conhecidos e não ao genoma completo, representando uma ordem de magnitude incorreta.

Mas acho que a resposta do outro cara ainda é muito engraçada.

Não me recordo exatamente do que respondi posteriormente — provavelmente algo como “**Não**” — mas essa ocasião em particular ficou gravada na minha memória por ter me proporcionado diversos *insights* sobre como as leis do pensamento são percebidas por aqueles que não estão esclarecidos.

Foi a primeira vez que me ocorreu que as intuições humanas faziam uma distinção qualitativa entre “Sem chance” e “Uma chance muito pequena, mas que vale a pena considerar”. Essa distinção ficou clara para mim durante um debate no blog *Overcoming Bias* (Superando o viés) sobre [loteria](#), quando alguém afirmou: “Existe uma grande diferença entre chance zero de ganhar e chance épsilon de ganhar”. Em resposta, eu disse: “Não, a diferença é da ordem de épsilon; se você tem dúvidas, deixe épsilon ser igual a um dividido por um Googleplex”.

O problema é que a teoria da probabilidade ocasionalmente nos permite calcular uma chance que é de fato tão pequena que é inútil dedicar espaço mental para acompanhá-la — mas, nesse ponto, você já a calculou. As pessoas confundem frequentemente o mapa com o território, levando a uma percepção visceral de que acompanhar uma probabilidade descrita simbolicamente é “uma chance que vale a pena”, mesmo que o referente numérico dessa descrição seja tão insignificante que, se fosse uma partícula de poeira, seria invisível aos nossos olhos. Podemos usar palavras para descrever números tão pequenos, mas não podemos usar palavras para descrever sentimentos tão diminutos. Um sentimento tão insignificante não existe, pois não dispara neurônios suficientes ou libera neurotransmissores em quantidade suficiente para ser sentido. É por isso que as pessoas compram bilhetes de loteria — ninguém pode realmente sentir a minúscula probabilidade envolvida.

No entanto, o que achei ainda mais fascinante foi a distinção qualitativa entre argumentos conside-

rados “certos” e “incertos”. Se um argumento não é considerado certo, podemos simplesmente ignorá-lo. Por exemplo, se a probabilidade de algo for zero, então devemos abandonar a crença. No entanto, se a probabilidade for algo como um sobre o googol ( $10^{100}$  ou, dez elevado a potência cem), podemos optar por mantê-la.

Agora vivemos em um país livre e ninguém deveria ser preso por raciocínio ilegal. No entanto, se estamos dispostos a ignorar um argumento que afirma uma probabilidade de um sobre googol, por que não ignorar também um argumento que afirma uma probabilidade de zero? Quero dizer, se já estamos ignorando as evidências de qualquer forma, por que seria considerado pior ignorar certas evidências em vez de evidências incertas?

Mas ainda há uma chance, certo?

Ao longo da vida, muitas vezes descobri que [aprendo](#) com exemplos negativos flagrantes de outras pessoas e aplico essas lições a situações mais sutis. Nesse contexto, a lição invertida é que se não podemos simplesmente ignorar uma probabilidade de um sobre googol apenas porque queremos, então também não podemos ignorar uma probabilidade de 0,9 apenas porque queremos. Ambos os casos representam abrir uma caixa de Pandora.

Considere o exemplo dele, caso você se pegue pensando: “Mas você não pode provar que estou errado”. Se você está disposto a ignorar um contra-argumento baseado em probabilidades, por que não ignorar também uma evidência comprovada?

## 51 - A falácia do cinza



O Sofisticado: “O mundo não é preto e branco. Ninguém é totalmente bom ou mau. Tudo é uma questão de tons de cinza. Portanto, ninguém é superior a ninguém.”

O Zetete: “Ao conhecer apenas o cinza, você conclui que todos os tons são iguais. Você ridiculariza a visão simplista de duas cores, mas a substitui por uma visão de cor única...”<sup>5</sup>

— Marc Stiegler, *David's Slings* [1] (O estilingue de Davi)

Não sei se o erro do Sofisticado tem um nome oficial, mas eu o chamo de Falácia do Cinza. Vimos essa falácia no ensaio anterior. Nele, alguém afirmava que as chances de dois elevado a 750 milhões contra um ainda significavam que “havia uma chance”. Para essa pessoa, todas as probabilidades eram simplesmente “incertas”. Isso significava que ela podia ignorá-las à vontade.

Por exemplo, “a Lua é feita de queijo verde” e “o Sol é composto principalmente de hidrogênio e hélio” são incertezas. Porém, não são equivalentes. Tudo são tons de cinza, mas alguns são tão claros que quase são brancos. Outros são tão escuros que quase são pretos. Mesmo assim, podemos comparar tons e dizer “este é mais escuro” ou “aquele é mais claro”.

Há alguns anos, vivi um momento marcante como racionalista. Foi quando li este trecho de *Player of Games* (O Jogador) de Iain M. Banks [2], especialmente a frase em negrito:

Um sistema que se julga culpado não reconhece inocentes. Isso acontece com qualquer poder que vê todos como aliados ou inimigos. Nós nos opomos a esse tipo de sistema. Você também se oporia, se refletisse sobre isso. Seu modo de pensar o coloca entre os inimigos deles. **Pode não ser sua culpa. Toda sociedade impõe alguns valores aos que nela crescem. Mas algumas sociedades tentam maximizar esse efeito, outras tentam minimizá-lo.** Você vem de uma das últimas e deve explicar para uma das primeiras. Dissimular será mais difícil do que imagina; a neutralidade é quase impossível. Você não pode escolher não ter suas visões políticas. Elas não são separáveis de você ou do resto do seu ser. São uma função da sua existência. Sei disso, eles sabem disso. É melhor aceitar<sup>6</sup>.

Considere um debate entre [Robin Hanson](#) e [Tyler Cowen](#). Hanson disse preferir dar 75% de peso às prescrições da teoria econômica, em vez de suas intuições: “Tento aplicar a teoria econômica diretamente, com pouco julgamento pessoal ou cultural.” Cowen respondeu:

---

5 NT: Texto original em inglês. *The Sophisticate*: “The world isn’t black and white. No one does pure good or pure bad. It’s all gray. Therefore, no one is better than anyone else”. *The Zetete*: “Knowing only gray, you conclude that all grays are the same shade. You mock the simplicity of the two-color view, yet you replace it with a one-color view. . .”

6 NT: Texto original em inglês. *A guilty system recognizes no innocents. As with any power apparatus which thinks everybody’s either for it or against it, we’re against it. You would be too, if you thought about it. The very way you think places you amongst its enemies. This might not be your fault, because every society imposes some of its values on those raised within it, but the point is that some societies try to maximize that effect, and some try to minimize it. You come from one of the latter and you’re being asked to explain yourself to one of the former. Prevarication will be more difficult than you might imagine; neutrality is probably impossible. You cannot choose not to have the politics you do; they are not some separate set of entities somehow detachable from the rest of your being; they are a function of your existence. I know that and they know that; you had better accept it.*

Na minha visão, não existe ‘teoria econômica de aplicação direta’. As teorias são sempre aplicadas através de nossos filtros pessoais e culturais. Não há outra maneira de ser.<sup>7</sup>

É possível minimizar esse efeito ou intensificá-lo. Em muitos casos, até na economia, não é irracional chamar o resultado de “direto” se nos esforçarmos para isso.

“Todo mundo é imperfeito”. Gandhi e Stalin eram ambos imperfeitos, mas de formas bem diferentes. “Toda pessoa é imperfeita” substitui uma visão de dois tons por uma de tom único. Dizer “ninguém é perfeito, mas alguns são menos imperfeitos que outros” pode não ser tão aplaudido. Mas há esperança para quem visa melhorar. Afinal, ninguém é perfeitamente imperfeito.

(Quando me dizem “O perfeccionismo é ruim”, respondo: “Ser imperfeito é aceitável, desde que não seja tão óbvio.”) Iguamente tola é a ideia de que “Todo paradigma científico impõe suas suposições na interpretação de experimentos”. Quem diz isso age como se provasse que a ciência está no mesmo nível da prática de curandeiros. Toda visão de mundo influencia nossas observações. Mas algumas tentam minimizar essa influência, enquanto outras a exaltam. Não existe só preto e branco. Há vários tons de cinza, uns mais claros que outros. É insensato tratá-los como iguais.

A Lua orbita a Terra há bilhões de anos. Você a vê no céu há anos e espera vê-la amanhã em sua fase prevista. Isso não é certeza absoluta. Esperar que um dragão invisível cure câncer também não é certeza. Mas são níveis de incerteza muito diferentes. Uma é a expectativa de que algo continue como sempre foi, previsto com precisão. A outra é a expectativa de algo que nunca foi observado.

Chamar ambos de “fé” é uma generalização excessiva. A ideia de que “a ciência também se baseia na fé” é estranha. Geralmente, quem diz isso vê a fé como algo positivo. Mas por que falam isso de forma irritada e triunfante, e não como elogio? É um argumento arriscado, se pensarmos bem. Se a ciência se baseia na “fé”, então ela seria comparável à religião.

Mas se a ciência é uma religião, é aquela que cura doenças e revela os segredos do universo. Poderíamos dizer: “Os sacerdotes da ciência caminham na Lua de forma pública e verificável. É um *milagre* baseado na fé científica, algo que a fé religiosa não consegue fazer.” Tem certeza de que quer seguir por esse caminho, ó, fiel? Talvez seja melhor repensar essa ideia de que “a ciência também é uma religião”. Observe esta dinâmica interessante: você tenta esclarecer algo confuso até que fique claro. De repente, alguém se ofende e protesta: “Mas não é preto no branco! É cinza!”. É diferente quando alguém diz: “É mais complexo do que parece, devido aos problemas X, Y e Z”. Outra coisa é gritar com raiva “Não é branco! É cinza!” sem explicar os pontos obscuros.

Nesse caso, suspeito de uma mentalidade problemática. É como se a pessoa tivesse feito as pazes com seus erros e agora rejeita qualquer chance de melhorar. Quem acha uma desculpa para não se esforçar frequentemente nega que outros possam fazê-lo. Cada forma de progresso se torna seu inimigo. Cada sugestão de avanço é vista como um insulto. Num momento, declaram com orgulho: “Fico feliz em ser cinza”. No seguinte, acusam com raiva: “Você também é cinza!”

Mesmo que não haja preto e branco puros, ainda existem tons mais claros e mais escuros. Nem todos os cinzas são iguais.

O comentarista [G2](#) nos remete ao ensaio de Asimov, *The Relativity of Wrong* (A Relatividade do Erro) [\[3\]](#):

Quando se acreditava que a Terra era plana, estava errado. Quando se pensava que a Terra era perfeitamente esférica, também estava errado. Mas se você acha que esses dois erros são equivalentes, sua visão está mais equivocada que ambas juntas.<sup>8</sup>

---

7 NT: Texto original em inglês. *In my view there is no such thing as “straightforwardly applying economic theory” . . . theories are always applied through our personal and cultural filters and there is no other way it can be.*

8 NT: Texto original em inglês. *When people thought the earth was flat, they were wrong. When people thought the earth was spherical, they were wrong. But if you think that thinking the earth is spherical is just as wrong as thinking the earth is flat, then your view is wronger than both of them put together.*

## Referências

- [1] Marc Stiegler, David's Sling (Baen, 1988).
- [2] Iain Banks, The Player of Games (Orbit, 1989).
- [3] Isaac Asimov, The Relativity of Wrong (Oxford University Press, 1989).



## 52 - Autoridade absoluta



Aquele que se aproxima de você com arrogância e diz: “A ciência realmente não sabe nada. Tudo o que você tem são teorias — você não pode ter certeza de estar certo. Vocês, cientistas, mudaram de ideia sobre como a gravidade funciona — quem pode dizer que amanhã vocês não mudarão de ideia sobre a evolução?”

Há uma imensa lacuna cultural aqui. Se você acha que pode atravessá-la em algumas frases, ficará muito desapontado.

No mundo dos não esclarecidos, existe autoridade e não autoridade. O que pode ser confiável, pode ser confiável; o que não é confiável, pode ser descartado. Existem fontes confiáveis de informação e fontes ruins de informação. Se os cientistas modificaram suas teorias ao longo da história, então a ciência não pode ser uma verdadeira autoridade e não pode mais ser considerada confiável — assim como uma testemunha pega em contradição ou um funcionário desonesto no caixa.

Além disso, essa pessoa presume que se espera que um defensor de uma ideia a defenda contra [todos os possíveis argumentos contrários](#) e não admita falhas. Todas as afirmações são desacreditadas de antemão. Se até mesmo os proponentes da ciência admitem que ela é imperfeita, então, ora, ela deve ser praticamente inútil.

Quando alguém viveu sua vida acostumado com certezas, não se pode simplesmente dizer a essa pessoa: “A ciência é probabilística, assim como todos os outros conhecimentos”. Ela aceitará a primeira parte da declaração como uma confissão de culpa e rejeitará a segunda parte como uma tentativa falha de culpar todos os outros para evitar o julgamento.

Ao admitir que não é totalmente confiável, a Ciência é dispensada e não deve mais nos incomodar!

Uma fonte óbvia desse padrão de pensamento é a religião, onde se afirma que as escrituras são provenientes de Deus; portanto, qualquer admissão de falhas nelas destruiria completamente sua autoridade. Assim, qualquer traço de dúvida é considerado pecado, e reivindicar certeza é obrigatório, independentemente de estar verdadeiramente convicto.

No entanto, suspeito que o sistema educacional tradicional também tenha sua parcela de culpa nesse aspecto. O professor lhe diz certas coisas, e você deve acreditar nelas e recitá-las na prova. Mas quando um aluno faz uma sugestão em sala de aula, você não é obrigado a concordar — você tem a liberdade de concordar ou discordar (ao que parece) e ninguém será punido por isso.

Receio que essa experiência esteja relacionada à crença nos domínios sociais da autoridade, do comando e da lei. No âmbito social, há uma diferença qualitativa entre leis absolutas e leis não absolutas, entre comandos e sugestões, entre autoridades e não autoridades. Parece haver um conhecimento rígido e um conhecimento flexível, assim como regulamentações estritas e regulamentações mais flexíveis. As autoridades estritas devem ser obedecidas, enquanto as sugestões mais flexíveis podem ser acatadas ou descartadas como uma questão de preferência pessoal. E a Ciência, uma vez que admite a possibilidade de erro, deve pertencer à segunda categoria.

(Notei, incidentalmente, que vejo certa semelhança com aqueles que acreditam que, se você não obtiver uma probabilidade autoritativa escrita em um pedaço de papel pelo professor em sala de aula, ou

transmitida por alguma Fonte Inquestionável similar, então sua incerteza não é um assunto para a teoria da probabilidade bayesiana. Alguém pode — suspirar! — contestar sua estimativa da probabilidade anterior. Portanto, para os não totalmente esclarecidos, parece que as *prioris* bayesianas pertencem à classe de crenças propostas pelos alunos e não à classe de crenças ditadas pelos professores — não é um conhecimento adequado.)

A lacuna cultural abissal entre o Caminho Autoritário e o Caminho Quantitativo é extremamente desconcertante para aqueles de nós que enxergam do ponto de vista racionalista. Aqui temos alguém que acredita possuir um conhecimento mais confiável do que as meras suposições probabilísticas da ciência — como a suposição de que a Lua surgirá em seu local e fase designados amanhã, exatamente como ocorre em todas as noites observadas desde a invenção do registro astronômico, e exatamente como previsto pelas teorias físicas cujas previsões anteriores foram confirmadas com sucesso em quatorze casas decimais. E qual é esse conhecimento que os não esclarecidos colocam acima do nosso, e por quê?

Provavelmente é algum pergaminho antigo e mofado que foi contradito de onze maneiras desde domingo, segunda-feira e todos os dias da semana. No entanto, isso é considerado mais confiável do que a Ciência (dizem eles) porque nunca admite erro, nunca muda de ideia, não importa quantas vezes seja refutado. Eles jogam a palavra “certeza” como uma bola de tênis, utilizando-a com tanta leviandade quanto uma pluma — enquanto os cientistas são consumidos pela dúvida, lutando para alcançar mesmo um mínimo de probabilidade. “Eu sou perfeito”, dizem eles, indiferentes ao mundo, “devo estar muito além de vocês, que continuam lutando para melhorar a si mesmos”.

Não há nada simples que se possa dizer a eles — nenhuma refutação rápida e esmagadora. Ao refletir cuidadosamente, é possível conquistar a audiência, se for um debate público. Infelizmente, não se pode simplesmente deixar escapar: “Tolo mortal, o Caminho Quantitativo está além de sua compreensão, e as crenças que você rotula levemente como ‘certas’ são menos seguras do que a menor de nossas poderosas hipóteses”. É uma diferença de visão de mundo que não pode ser facilmente descrita em palavras, muito menos de forma rápida.

O que se pode tentar, retoricamente, diante de uma plateia? É difícil dizer... talvez:

- “O poder da ciência reside na capacidade de mudarmos de opinião e admitir que estamos errados. Se você nunca admite estar errado, isso não significa que comete menos erros”.
- “Qualquer um pode afirmar estar absolutamente certo. É um pouco mais desafiador nunca cometer erros. Os cientistas entendem essa diferença, por isso não afirmam ter certeza absoluta. Isso é tudo. Isso não significa que eles tenham qualquer motivo específico para duvidar de uma teoria — absolutamente qualquer fragmento de evidência pode estar apontando na mesma direção, todas as estrelas e planetas alinhados como dominós em apoio a uma única hipótese, e os cientistas ainda não afirmarão que têm certeza, porque possuem padrões mais elevados. Isso não significa que os cientistas tenham menos direito à certeza do que, digamos, os políticos que sempre parecem ter tanta certeza de tudo”.
- “Os cientistas não utilizam a frase “não tenho certeza absoluta” da maneira que você está acostumado em uma conversa comum. Quero dizer, imagine que você vá ao médico e faça um exame de sangue, e o médico volte e diga: ‘Realizamos alguns testes e não há certeza absoluta de que você não seja feito de queijo, e há uma chance não-nula de que vinte fadas feitas de chocolate senciente estejam cantando a música ‘eu te amo’ do Barney dentro do seu intestino inferior’. Nesse caso, é melhor correr para as colinas, pois o seu médico precisa de um médico. Quando um cientista diz a mesma coisa, significa que eles consideram a probabilidade tão pequena que não pode ser vista nem mesmo com um microscópio eletrônico, mas o cientista está disposto a considerar a evidência no caso extremamente improvável de que ela exista.”
- “Você estaria disposto a mudar de opinião sobre as coisas que chama de ‘certas’ se visse evidências suficientes? Quero dizer, imagine que o próprio Deus descesse das nuvens e dissesse que toda a sua religião é verdadeira, exceto pelo nascimento virginal. Se isso mudasse sua mente, então você não pode afirmar que está absolutamente certo sobre o nascimento virginal. Por razões técnicas da teoria da probabilidade, se é teoricamente possível que você mude de opinião sobre algo, então não pode atribuir uma probabilidade exatamente igual a um. A incerteza pode ser menor do que uma partícula de poeira, mas precisa estar presente. E se você não mudaria de opinião mesmo que Deus lhe dissesse-

se o contrário, então você tem um problema em se recusar a admitir que está errado, e isso vai além de qualquer coisa que um mortal como eu possa lhe dizer, eu acredito.”

Porém, de certa forma, a questão mais interessante é como você aborda alguém que não está diante de uma plateia. Como você inicia o longo processo de ensinar alguém a viver em um universo sem certezas?

Acredito que o primeiro passo deve ser compreender que é possível viver sem certezas — isto é, mesmo que hipoteticamente não seja possível ter certeza de nada, isso não nos priva da capacidade de fazer distinções morais ou factuais. Parafraçando Lois Bujold, “Não empurre com mais força, diminua a resistência”.

Uma das defesas comuns da Autoridade Absoluta é algo que chamo de “O Argumento do Argumento Cinza”, que funciona da seguinte maneira:

Os relativistas morais dizem:

- “O mundo não é preto e branco, portanto:
- Tudo é cinza, portanto:
- Ninguém é melhor do que ninguém, então:
- Posso fazer o que eu quiser, e você não pode me impedir “muahahaha”<sup>9</sup>.
- No entanto, é imprescindível que consigamos evitar que as pessoas cometam assassinatos.
- Portanto, deve haver alguma maneira de estar absolutamente certo, ou os relativistas morais vencem.

[Estupidez ao contrário](#) não é inteligência. Você não pode chegar a uma resposta correta invertendo cada linha de um argumento que chega a uma conclusão ruim — isso daria ao tolo muito controle sobre você. Cada linha deve estar correta para um argumento matemático ser válido. O fato de os relativistas morais dizerem que “O mundo não é preto e branco” não faz com que isso seja falso. Não decorre da crença de Stalin de que  $2 + 2 = 4$ , que “ $2 + 2 = 4$ ” seja falso. O erro (e precisa apenas um) está no salto de uma perspectiva de duas cores para uma perspectiva de somente uma cor, para a ideia de que todos os cinzas têm o mesmo tom.”

Seria conceder demais (na verdade, conceder todo o argumento) concordar com a premissa de que é necessário ter conhecimento absoluto sobre opções absolutamente boas e opções absolutamente más para ser moral. É possível ter um conhecimento incerto sobre opções relativamente melhores e relativamente piores e ainda assim fazer uma escolha. Deveria ser algo rotineiro, na verdade, e não algo dramático.

Quero dizer, sim, se você estiver diante de duas alternativas, A e B, e de alguma forma conseguir estabelecer uma confiança de 100%, sabidamente bem calibrada, de que A é absoluta e inteiramente desejável, enquanto B é a soma de tudo o que é mau e repugnante, então isso é uma condição suficiente para escolher A em vez de B. Mas não é uma condição necessária.

Ah, e vale mencionar: falácia lógica: apelo às consequências da crença.

Vejamos, o que mais eles precisam saber? Bem, existe toda uma cultura racionalista que afirma que a dúvida, o questionamento e a confissão do erro não são coisas terrivelmente vergonhosas.

Existe também a ideia de obter informações ao observar as coisas, em vez de fazer proselitismo. Quando você olha para as coisas com mais atenção, às vezes descobre que elas são diferentes do que você pensava à primeira vista. Mas isso não significa que a natureza lhe mentiu ou que você deva desistir de enxergar. Além disso, há o conceito de confiança calibrada: essa “probabilidade” não é o mesmo conceito que a pequena barra de progresso em sua mente que mede seu comprometimento emocional com uma ideia. É mais uma medida de quantas vezes, pragmaticamente, na vida real, as pessoas em um certo estado de crença dizem coisas que são realmente verdadeiras. Se você pegar cem pessoas e pedir a cada uma delas para fazer uma afirmação da qual estão “absolutamente certas”, quantas dessas afirmações estarão corretas? Certamente não serão cem.

Na verdade, as afirmações pelas quais as pessoas são verdadeiramente fanáticas têm muito menos

---

9 NT: Onomatopeia utilizada em português brasileiro para representar uma risada maligna. No original em inglês o autor utiliza a onomatopeia *bwahahaha* com a mesma intenção.

probabilidade de estarem corretas do que afirmações como “o Sol é maior que a Lua”, que parecem tão óbvias que não geram entusiasmo. Para cada afirmação em que alguém está “absolutamente certo”, provavelmente você encontrará alguém “absolutamente certo” do contrário, porque tais profissões fanáticas de crença não surgem na ausência de oposição. Portanto, a pequena barra de progresso na mente das pessoas, que mede seu compromisso emocional com uma crença, não se traduz bem em uma confiança calibrada – e nem mesmo se comporta consistentemente.

Quanto à “certeza absoluta” — bem, se você diz que algo é 99,9999% provável, significa que você acredita que poderia fazer um milhão de declarações independentes igualmente fortes, uma após a outra, ao longo de um ano inteiro ou mais, e estar errado, em média, cerca de uma vez. Isso é impressionante o suficiente. (É impressionante perceber que podemos realmente alcançar esse nível de confiança ao dizer “[Você não ganhará na loteria](#)”.) Portanto, não vamos mencionar probabilidades de 1,0. Após perceber que não precisamos de probabilidades de 1,0 para ter sucesso na vida, você perceberá como é absolutamente ridículo pensar que poderíamos chegar a 1,0 com um cérebro humano. Uma probabilidade de 1,0 não é apenas certeza, é certeza infinita.

Na verdade, parece-me que, para evitar mal-entendidos públicos, talvez os cientistas devessem dizer “Não temos certeza infinita” em vez de “Não temos certeza”. Porque esta última frase, no discurso comum, sugere que você conhece algum motivo específico para ter dúvida.

## 53 - Como me convencer de que $2 + 2 = 3$



Em “O que são evidências?”, escrevi:

É por isso que os racionalistas atribuem tanto valor à alegação aparentemente paradoxal de que uma crença só é verdadeiramente válida se você puder, em princípio, ser convencido do contrário. Se a sua retina permanecesse no mesmo estado, independentemente da luz que entrasse, você ficaria cego... Daí o ditado “fé cega”. Se o que você acredita não depende do que você vê, você foi tão cegamente iludido quanto se tivesse fechado os olhos.

Cihan Baran [respondeu](#):

Eu não consigo conceber uma situação em que  $2 + 2$  igual a 4 seja falso. Talvez seja por isso que minha crença em  $2 + 2 = 4$ , seja incondicional<sup>10</sup>.

Admito que não consigo conceber uma “situação” em que  $2 + 2$  seja falso. (Existem redefinições, mas elas não são “situações” e, nesse caso, você não estaria mais falando sobre 2, 4, = ou +). No entanto, isso não torna minha crença incondicional. É perfeitamente possível para mim imaginar uma situação em que eu possa ser convencido de que  $2 + 2 = 3$ .

Vamos supor que eu acorde de manhã, pegue dois tampões de ouvido e os coloque ao lado de outros dois tampões de ouvido na minha mesa de cabeceira, e perceba que agora existem três tampões de ouvido, sem que nenhum deles tenha aparecido ou desaparecido — em contraste com minha memória que diz que  $2 + 2$  deveria ser igual a 4.

Além disso, quando visualizei o processo em minha mente, parecia que fazer  $xx$  e  $xx$  resultar em  $xxxx$  exigia um  $x$  adicional que surgia do nada, o que era inconsistente com outras operações aritméticas que visualizei, pois subtrair  $xx$  de  $xxx$  resultava em  $xx$ , enquanto subtrair  $xx$  de  $xxxx$  resultava em  $xxx$ . Isso entraria em conflito com minha memória que diz que  $3 - 2 = 1$ , mas essa memória seria absurda diante da confirmação física e mental de que  $xxx - xx = xx$ .

Eu também conferiria com uma calculadora de bolso, o Google e até mesmo minha cópia de 1984<sup>11</sup>, onde Winston escreve que “Liberdade é a liberdade de dizer que dois mais dois são três”. Tudo isso, naturalmente, demonstraria que o restante do mundo concordava com minha interpretação atual e discordava da minha memória, segundo a qual  $2 + 2$  é igual a 3.

Como pude ser tão enganado a ponto de acreditar que  $2 + 2 = 4$ ? Duas explicações vêm à mente: primeira, uma falha neurológica (possivelmente causada por um espirro) fez com que todas as somas aditivas na minha memória armazenada aumentassem em um. Segunda, alguém estava brincando comigo, seja por meio de hipnose ou porque sou uma simulação de computador. No segundo caso, acredito ser mais provável que tenham mexido com a minha memória aritmética do que  $2 + 2$  ser realmente igual a 4. Nenhuma dessas explicações aparentemente plausíveis impediria que eu percebesse estar extremamente confuso.

O que me convenceria de que  $2 + 2 = 3$ , em outras palavras, é o mesmo tipo de evidência que atualmente me convence de que  $2 + 2 = 4$ : uma convergência de evidências provenientes da observação física,

10 NT: Texto original em inglês. *I can not conceive of a situation that would make  $2 + 2 = 4$  false. Perhaps for that reason, my belief in  $2 + 2 = 4$  is unconditional.*

11 NT: Obra clássica de George Orwell.

visualização mental e acordo social. Houve um tempo em que eu não tinha ideia de que  $2 + 2 = 4$ . Não cheguei a essa nova crença por processos aleatórios, então não haveria motivo específico para o meu cérebro armazenar “ $2 + 2 = 4$ ” em vez de “ $2 + 2 = 7$ ”. O fato de o meu cérebro armazenar uma resposta surpreendentemente similar ao que ocorre quando coloco dois tampões de ouvido ao lado de outros dois tampões de ouvido sugere fortemente uma explicação de que o emaranhamento produz esse estranho reflexo entre mente e realidade.

Na verdade, existem apenas duas possibilidades para uma crença ser verdadeira — ou a crença surgiu por meio de um processo de emaranhamento mente-realidade ou não. Caso contrário, a crença não pode estar correta, a menos por coincidência. Para crenças com o menor grau de complexidade interna (requerendo um software de mais de 10 bits para simular), o espaço de possibilidades é grande o suficiente para que a coincidência se torne insignificante.

Fatos incondicionais não devem ser confundidos com crenças incondicionais. Se evidências emaranhadas me convencem de que um fato é incondicional, isso não significa que sempre acreditei nesse fato sem a necessidade de evidências emaranhadas.

Acredito que  $2 + 2 = 4$  e acho muito fácil conceber uma situação que me convenceria de que  $2 + 2 = 3$ . Ou seja, seria o mesmo tipo de situação que atualmente me convence de que  $2 + 2 = 4$ . Portanto, não tenho medo de ser vítima de uma fé cega.

Se houver algum cristão na plateia que esteja familiarizado com o Teorema de Bayes (sem fobias numéricas, por favor), posso perguntar a você que tipo de situação o convenceria da verdade do Islã? Presumivelmente, seria o mesmo tipo de situação que causalmente levou à sua crença atual no cristianismo: você seria empurrado aos berros para fora do útero de uma mulher muçulmana e seria criado por pais muçulmanos que constantemente lhe diriam que acreditar incondicionalmente no Islã é bom. Ou há algo mais, além disso? Se sim, que tipo de situação o convenceria do Islã ou, pelo menos, de abandonar o cristianismo?

## 54 - Certeza Infinita



Em "[Autoridade absoluta](#)", argumentei que você não precisa de certeza infinita:

Caso precise escolher entre as alternativas A e B, e consiga estabelecer, de alguma forma, uma confiança plenamente calibrada de que A é absolutamente desejável e que B representa tudo o que é ruim e repugnante, então esse argumento é suficiente para tomar a decisão em favor de A em detrimento de B. No entanto, essa condição não é necessária. É possível ter um conhecimento incerto sobre opções relativamente melhores ou piores e ainda assim fazer uma escolha. Na verdade, essa deveria ser uma prática comum.

No que diz respeito à afirmação de que  $2 + 2 = 4$ , é importante distinguir entre a representação e a realidade. Considerando a aparente estabilidade absoluta e universalidade das leis físicas, é plausível que, ao longo de toda a história do universo, nenhuma partícula jamais tenha ultrapassado a velocidade local da luz. Isso significa que o limite de velocidade da luz pode não ser apenas verdadeiro, 99% do tempo, ou 99,9999% do tempo, ou mesmo  $(1 - 1/\text{googolplex})$  do tempo, mas sim sempre e absolutamente verdadeiro.

No entanto, a questão de ter confiança absoluta no limite da velocidade da luz é outra questão. O mapa não é o território.

Pode ser totalmente verdade que um aluno tenha plagiado sua tarefa, mas ter algum conhecimento desse fato — ou ainda menos confiança absoluta nessa crença — é uma questão separada. Se você jogar uma moeda e não olhar para ela, pode ser completamente verdade que a moeda está mostrando cara, mas você pode não ter certeza se ela está mostrando cara ou coroa. Um grau de incerteza não é o mesmo que um grau de verdade ou uma frequência de ocorrência.

O mesmo se aplica às verdades matemáticas. É discutível se a afirmação " $2 + 2 = 4$ " ou "Na aritmética de Peano,  $SS0 + SS0 = SSSS0$ " pode ser considerada verdadeira em qualquer sentido puramente abstrato, além de sistemas físicos que parecem se comportar de maneira semelhante aos axiomas de Peano. Dito isso, vou avançar e supor que, em qualquer sentido em que " $2 + 2 = 4$ " seja verdadeiro, ele é sempre e precisamente verdadeiro, não apenas aproximadamente verdadeiro (" $2 + 2$  é realmente igual a 4,0000004") ou verdadeiro 999.999.999.999 vezes em 1.000.000.000.000.

Não tenho certeza do que "verdadeiro" deve significar nesse caso, mas sustento minha suposição. A credibilidade de " $2 + 2 = 4$  é sempre verdadeiro" supera em muito a credibilidade de qualquer posição filosófica específica sobre o significado de "verdadeiro", "sempre" ou "é" na declaração acima.

Porém, isso não significa que eu tenha confiança absoluta de que  $2 + 2 = 4$ . Veja a discussão anterior sobre [como me convencer de que  \$2 + 2 = 3\$](#) , o que poderia ser feito usando o mesmo tipo de evidência que inicialmente me convenceu de que  $2 + 2 = 4$ . Pode ser que eu tenha imaginado todas as evidências anteriores ou que eu esteja me lembrando erroneamente. Existem disfunções cerebrais registradas na neurologia que são ainda mais estranhas do que essa.

Portanto, se atribuirmos alguma probabilidade à afirmação " $2 + 2 = 4$ ", qual deveria ser essa probabilidade? O que se busca alcançar nesse caso é uma boa calibração — declarações às quais se atribui "99% de probabilidade" se concretizam 99 vezes a cada 100. Na verdade, isso é muito mais difícil do que se imagina. Pegue cem pessoas e peça a cada uma delas para fazer dez afirmações sobre as quais elas tenham "99% de confiança". Das 1.000 afirmações, você acredita que cerca de 10 estarão erradas?

Não vou discutir os experimentos reais realizados em calibração — você pode encontrá-los no capítulo

lo do meu livro intitulado [Cognitive biases potentially affecting judgment of global risks](#) (Vieses cognitivos que potencialmente afetam a avaliação dos riscos globais) — porque, quando menciono isso casualmente para pessoas sem a devida preparação, elas o utilizam como um [contra-argumento bastante generalizado](#), que surge em suas mentes sempre que desejam questionar a confiança de alguém cuja opinião não apreciam, mas não está disponível quando se trata de suas próprias opiniões. Portanto, tento não discutir os experimentos de calibração, a menos que seja como parte de uma apresentação estruturada sobre racionalidade que inclua advertências contra o ceticismo motivado.

Mas a calibração observada em seres humanos que afirmam estar “99% confiantes” não atinge uma precisão de 99%.

Suponhamos que você afirme ter 99,99% de confiança de que  $2 + 2 = 4$ . Ao fazer isso, você está basicamente afirmando que poderia fazer 10.000 declarações independentes, depositando igual confiança em cada uma delas, e cometer um erro em média apenas uma vez. Talvez, para a igualdade  $2 + 2 = 4$ , esse grau extraordinário de confiança seja possível: é uma afirmação extremamente simples, matemática, empírica e amplamente aceita socialmente (não como uma afirmação apaixonada, mas como algo discretamente aceito como verdadeiro). Então, talvez seja realmente possível alcançar uma confiança de até 99,99% nesse caso.

Duvido que seja possível alcançar uma confiança de 99,99% para afirmações como “53 é um número primo”. Sim, isso parece provável, mas quando você tenta estabelecer protocolos que permitam fazer 10.000 declarações independentes desse tipo — ou seja, não apenas um conjunto de declarações sobre números primos, mas um novo protocolo a cada vez — você falhará mais de uma vez. Peter de Blanc tem uma anedota engraçada sobre esse ponto. (Eu disse a ele para não fazer isso novamente.)

No entanto, o mapa não é o território: quando digo que tenho 99% de confiança de que  $2 + 2 = 4$ , isso não significa que acredito que “ $2 + 2 = 4$ ” seja verdade com uma precisão de 99% ou que seja verdade em 99 de 100 vezes. A proposição na qual deposito minha confiança é a de que “ $2 + 2 = 4$  é sempre e exatamente verdadeiro”, não a de que “ $2 + 2 = 4$  é principal e geralmente verdadeiro”.

Quanto à ideia de alcançar 100% de confiança em uma proposição matemática — bem, vamos lá! Se você disser que tem 99,9999% de confiança, está sugerindo que poderia fazer um milhão de declarações igualmente carregadas, uma após a outra, e estar errado, em média, apenas uma vez. Isso significaria falar constantemente por um período aproximado de um ano, se você fizesse uma afirmação a cada 20 segundos e falasse 16 horas por dia.

Agora, se você afirmar 99,9999999999% de confiança, você estaria chegando a um trilhão. Nesse caso, você teria que falar por cem vidas humanas sem cometer um único erro.

Se você afirmar uma confiança de  $(1 - 1/\text{googolplex})$  (um sobre um googolplex), seu ego estará além daquele de pacientes mentais que pensam ser Deus.

E um googolplex é consideravelmente menor do que [números relativamente pequenos e inconcebivelmente grandes](#), como  $3 \uparrow \uparrow \uparrow 3$  (três seta tripla três). No entanto, mesmo com uma confiança de  $(1 - 1/3 \uparrow \uparrow \uparrow 3)$  (Um para um sobre três seta tripla três), ainda não estamos muito mais perto de uma **PROBABILIDADE 1** do que de ter 90% de certeza sobre algo. Se tudo mais falhar, os hipotéticos Lordes das Trevas da Matrix, que atualmente estão distorcendo a avaliação de credibilidade do seu cérebro para esta mesma frase, irão bloquear o caminho e nos proteger da maldição da certeza infinita.

Estou absolutamente certo disso? Claro que não.

Como Rafal Smigrodski disse uma vez:

Eu diria que é possível atribuir um nível de certeza inferior a 1 aos conceitos matemáticos necessários para derivar a própria regra de Bayes e ainda assim utilizá-la de maneira prática. Não posso ter certeza absoluta, devo sempre permanecer inseguro. Talvez haja situações em que eu possa ter uma certeza legítima sobre algo. No entanto, uma vez que atribuo uma probabilidade de 1 a uma proposição, nunca mais posso revertê-la. Não importa o que eu veja ou aprenda, sou obrigado a rejeitar qualquer coisa que contradiga o



axioma. Não gosto da ideia de não poder mudar de ideia, nunca.<sup>12</sup>

---

12 NT.: Texto original em inglês. *I would say you should be able to assign a less than 1 certainty level to the mathematical concepts which are necessary to derive Bayes's rule itself, and still practically use it. I am not totally sure I have to be always unsure. Maybe I could be legitimately sure about something. But once I assign a probability of 1 to a proposition, I can never undo it. No matter what I see or learn, I have to reject everything that disagrees with the axiom. I don't like the idea of not being able to change my mind, ever.*

## 55 - 0 e 1 não são probabilidades



Um, dois e três são todos números inteiros, assim como o número negativo quatro. Se você continuar contando para cima ou para baixo, certamente encontrará uma infinidade de números inteiros. No entanto, você não encontrará nada chamado de “infinito positivo” ou “infinito negativo”, pois eles não são considerados números inteiros.

O infinito positivo e negativo não são números inteiros, mas sim símbolos especiais usados para descrever o comportamento dos números inteiros. Às vezes, as pessoas dizem coisas como “ $5 + \text{infinito} = \text{infinito}$ ”, porque se você começar com 5 e continuar contando indefinidamente, obterá números crescentes sem um limite. No entanto, não podemos concluir que “ $\text{infinito} - \text{infinito} = 5$ ”. Não é possível contar a partir de 0 sem parar e depois contar regressivamente sem parar e, ao final, chegar a 5.

A partir disso, podemos perceber que o infinito não é apenas um não inteiro, mas também não se comporta como um número inteiro. Se você tentar misturar infinito com números inteiros imprudentemente, acabará com comportamentos aparentemente inconsistentes que não são observados nos números reais como 1, 2, 3 e outros.

Mesmo que o infinito não seja um número inteiro, não precisamos nos preocupar em ficar sem números. Embora as pessoas já tenham visto cinco ovelhas, milhões de grãos de areia e septilhões de átomos, ninguém jamais contou uma quantidade infinita de algo. O mesmo vale para quantidades contínuas — as pessoas mediram partículas de poeira com um milímetro de diâmetro, animais com um metro de altura, cidades com quilômetros de extensão e galáxias com milhares de anos-luz de diâmetro, mas ninguém jamais mediu nada com um diâmetro infinito. No mundo real, não precisamos de infinito em grande escala.

Nota para leitores mais sofisticados: Não é necessário me escrever com explicações detalhadas sobre, por exemplo, a diferença entre números ordinais e cardinais. Embora eu esteja familiarizado com várias definições avançadas de conjuntos teóricos infinitos, não vejo uma aplicação útil para eles na teoria da probabilidade.

Na forma convencional de expressar probabilidades, estas estão compreendidas entre 0 e 1. Por exemplo, uma moeda pode ter uma probabilidade de 0,5 de cair em coroa, ou um meteorologista pode atribuir uma probabilidade de 0,9 para chuva amanhã.

No entanto, essa não é a única forma de expressar probabilidades. Por exemplo, é possível transformar probabilidades em chances por meio da fórmula  $O = (P / (1 - P))$ . Dessa forma, uma probabilidade de 50% seria transformada em chances de 0,5/0,5 ou 1, geralmente representadas como 1:1, enquanto uma probabilidade de 0,9 seria transformada em chances de 0,9/0,1 ou 9, geralmente escritas como 9:1. Para converter chances em probabilidades, utiliza-se a fórmula  $P = (O / (1 + O))$ , e essa conversão é perfeitamente reversível, tornando a transformação um isomorfismo — um mapeamento bidirecional reversível. Portanto, probabilidades e chances são isomórficas e pode-se utilizar uma ou outra conforme a conveniência.

Por exemplo, é mais conveniente utilizar probabilidades ao realizar atualizações bayesianas. Suponha que eu lance um dado de seis faces: se qualquer face, exceto o número 1, aparecer, há uma chance de 10% de ouvir um sino, mas se a face 1 aparecer, a chance de ouvir o sino é de 20%. Agora, suponha que eu lance o dado e ouça o sino. Qual é a chance de que o número mostrado seja 1? Bem, as probabilidades iniciais são de 1:5 (correspondendo ao número real  $1/5 = 0,20$ ) e a razão de verossimilhança é de 0,2:0,1 (correspondendo ao número real 2). Multiplicando esses dois valores, obtemos as probabilidades posteriores de 2:5 (corres-

pondo ao número real  $2/5$  ou  $0,40$ ). Em seguida, se eu quiser, posso converter novamente para uma probabilidade, obtendo  $(0,4/1,4) = 2/7 = \sim 29\%$ .

Portanto, as probabilidades são mais manejáveis para atualizações bayesianas — se você utilizar probabilidades, precisará aplicar o Teorema de Bayes em sua forma mais complicada. No entanto, as probabilidades são mais convenientes para responder a perguntas como “Se eu lançar um dado de seis lados, qual é a chance de obter um número de 1 a 4?” Nesse caso, é possível somar as probabilidades de  $1/6$  para cada lado e obter  $4/6$ , mas não é possível somar as razões de chances de  $0,2$  para cada lado e obter uma razão de chances de  $0,8$ .

Por que estou compartilhando tudo isso? Para mostrar que as razões de chances são uma forma tão válida de mapear incertezas em números reais quanto as probabilidades. As razões de chances são mais convenientes para certas operações, enquanto as probabilidades são mais convenientes para outras. Um famoso teorema chamado Teorema de Cox (juntamente com suas diversas extensões e refinamentos) demonstra que todas as formas de representar incertezas que obedecem a algumas restrições aparentemente razoáveis são isomórficas entre si. Por que é importante afirmar que as razões de chances são tão legítimas quanto as probabilidades?

As probabilidades, normalmente expressas entre 0 e 1, parecem ser quantidades facilmente compreensíveis — é fácil visualizar 1 zebra ou 0 unicórnios. No entanto, quando você transforma as probabilidades em razões de chances, o valor 0 permanece o mesmo, mas o valor 1 é representado como infinito positivo. Agora, a noção de verdade absoluta não parece tão acessível.

Uma representação que simplifica ainda mais as atualizações bayesianas, é o *log odds* (ou logaritmo das chances) — é assim que E. T. Jaynes recomendava pensar sobre probabilidades. Por exemplo, suponha que a probabilidade inicial de uma proposição seja  $0,0001$  — isso corresponde a um log de probabilidades de aproximadamente  $-40$  decibéis. Em seguida, você observa evidências que parecem ser 100 vezes mais prováveis se a proposição for verdadeira em comparação com ser falsa, o que equivale a um acréscimo de 20 decibéis de evidência. Portanto, as probabilidades posteriores ficam em torno de  $-40 \text{ dB} + 20 \text{ dB} = -20 \text{ dB}$ , ou seja, a probabilidade posterior é de aproximadamente  $\sim 0,01$ .

Quando você converte probabilidades em log odds, o valor 0 se transforma em infinito negativo, enquanto o valor 1 se transforma em infinito positivo. Agora, tanto a certeza absoluta quanto a improbabilidade absoluta parecem um pouco mais inatingíveis.

Em probabilidades tradicionais,  $0,9999$  e  $0,99999$  parecem estar separados por apenas  $0,00009$ , enquanto  $0,502$  está muito mais distante de  $0,503$  do que  $0,9999$  está de  $0,99999$ . Parece que para chegar da probabilidade  $0,999$  à probabilidade 1, você só precisa percorrer uma distância de  $0,00001$ .

No entanto, quando você converte para razão de chances,  $0,502$  e  $0,503$  se transformam em  $1,008$  e  $1,012$ , respectivamente, enquanto  $0,9999$  e  $0,99999$  se transformam em  $9.999$  e  $99.999$ . E quando você converte para log odds,  $0,502$  e  $0,503$  se tornam  $0,03$  decibéis e  $0,05$  decibéis, mas  $0,9999$  e  $0,99999$  se tornam  $40$  decibéis e  $50$  decibéis.

Ao trabalhar com probabilidades logarítmicas, **a distância entre quaisquer dois níveis de incerteza é igual à quantidade de evidência necessária para passar de um para o outro**. Em outras palavras, as probabilidades logarítmicas nos fornecem uma medida natural do espaçamento entre os diferentes graus de confiança.

O uso de probabilidades logarítmicas revela que alcançar a [certeza absoluta](#) requer evidências infinitamente fortes, assim como a improbabilidade absoluta requer evidências contrárias infinitamente fortes.

Além disso, os teoremas padrão em probabilidade possuem casos especiais quando você tenta inserir 1s (uns) ou 0s (zeros) neles, como o que acontece ao realizar uma atualização bayesiana em uma observação à qual foi atribuída uma probabilidade de 0.

Portanto, proponho que seja sensato afirmar que 1 e 0 não fazem parte das probabilidades, assim como o infinito negativo e positivo, que não obedecem aos axiomas de campo, não estão presentes nos números reais.

A principal preocupação dos teóricos da probabilidade em relação a essa afirmação é que seria necessário derivar novamente os teoremas obtidos anteriormente, assumindo que podemos obter probabilidades marginais somando todas as partes e igualando-as a 1. No entanto, no mundo real, quando você joga um dado, ele não tem uma certeza infinita de cair em algum número entre 1 e 6. O dado pode ficar em equilíbrio na borda, ser atingido por um meteoro ou os Senhores das Trevas da Matrix podem entrar e escrever “37” em uma das faces.

Se criássemos um símbolo mágico para representar “todas as possibilidades que não foram consideradas”, então poderíamos marginalizar os eventos, incluindo esse símbolo mágico, e chegar a um símbolo mágico “T” que representaria certeza infinita.

No entanto, prefiro questionar se existe uma maneira de derivar um teorema sem o uso de símbolos mágicos com comportamentos especiais. Isso seria mais elegante. Assim como existem matemáticos que se recusam a acreditar na lei dos conjuntos intermediários excluídos ou no infinito, gostaria de ser um teórico da probabilidade que não acredita na certeza absoluta.

## 56 - Sua Racionalidade é da Minha Conta



Algumas respostas à [“Loterias: uma perda de esperança”](#). Fui repreendido por ousar criticar as decisões alheias pela obra *A Waste of Hope* (Desperdício de Esperança). Se alguém opta por comprar bilhetes de loteria, quem sou eu para discordar? Isso é apenas um caso específico de uma questão mais ampla: por que devo me importar se alguém prefere acreditar no que é agradável em vez do que é verdadeiro? Não podemos todos escolher, individualmente, se nos importamos com a verdade?

Uma resposta mal-humorada óbvia seria: “Por que você se importa se eu me importo com o fato de alguém se importar com a verdade?” É um tanto inconsistente ter um termo negativo em sua função de utilidade enquanto se tem um termo para a função de utilidade de outra pessoa. Mas isso é apenas uma observação rápida, não uma resposta.

Portanto, aqui está a minha resposta: acredito que seja correto e adequado para mim, enquanto ser humano, demonstrar interesse pelo futuro e pelo rumo que a civilização humana tomará. Um desses interesses é a busca da verdade, a qual tem sido gradualmente fortalecida ao longo das gerações (pois nem sempre existiu a ciência). Desejo promover ainda mais essa busca nesta geração. Esse é o meu desejo para o futuro. Afinal, todos somos participantes nesse vasto jogo, independentemente de aceitarmos essa responsabilidade ou não.

E isso faz com que sua racionalidade me diga respeito.

Essa é uma ideia perigosa? Sim, e não apenas de maneira superficialmente preocupante. Houve casos em que pessoas foram queimadas vivas porque algum líder religioso decidiu que elas não pensavam da maneira que deveriam. A noção de queimar alguém vivo simplesmente porque ele “pensa errado” é um tipo de raciocínio revoltante, não é? É repugnante pensar que as pessoas poderiam ter tal mentalidade. Aqueles que pensam assim exigem uma ação.

Concordo plenamente! Aqui está a minha proposta: argumentemos contra ideias prejudiciais, mas sem recorrer à violência.

O silogismo que devemos evitar é o seguinte: “Acredito que Susie expressou uma ideia ruim; portanto, Susie deve ser queimada viva.” Alguns tentam evitar esse silogismo rotulando-o como inapropriado acreditar que Susie tenha dito algo errado. Eles acreditam que ninguém deve julgar ninguém, em nenhum momento; qualquer pessoa que faça um julgamento está cometendo um pecado terrível e deve ser publicamente denunciada por isso.

No entanto, em relação a mim, rejeito esse “portanto”. Meu silogismo é o seguinte: “Acredito que Susie expressou algo errado; portanto, vou argumentar contra o que ela disse, mas não vou queimá-la viva nem tentar silenciá-la por meio de violência ou regulamentação...”

Somos todos participantes nesse vasto tabuleiro, e um dos meus interesses para o futuro é tornar o jogo mais justo. A ideia contraintuitiva subjacente à ciência é que desacordos factuais devem ser combatidos por meio de experimentos e matemática, não por meio de violência e decretos. Essa incrível noção pode ser estendida além da ciência, para uma luta justa por todo o futuro. A vitória deve ser alcançada convencendo as pessoas, e não por meio de violência. Esse é um dos princípios da racionalidade, ao qual prometi minha lealdade.

As pessoas que defendem o relativismo ou o egoísmo não me parecem verdadeiramente relativistas,

ou egoístas. Se fossem verdadeiramente relativistas, não julgariam. Se fossem verdadeiramente egoístas, estariam focadas apenas em ganhar dinheiro em vez de se envolverem em discussões acaloradas com os outros. Em vez disso, escolheram o lado do relativismo, cujo objetivo nesse vasto tabuleiro é impedir que os participantes — todos os participantes — façam certos tipos de julgamentos. Ou escolheram o lado do egoísmo, cujo objetivo é tornar todos os participantes egoístas. E assim, eles jogam o jogo, de forma justa ou injusta, de acordo com sua sabedoria.

Se existem verdadeiros relativistas ou egoístas, não os ouvimos — eles permanecem em silêncio, não são jogadores ativos.

Não posso deixar de me importar com o que você pensa, pois assim como não consigo evitar ver o universo, a cada vez que um ser humano se distancia da verdade, o curso da história da humanidade se torna um pouco mais sombrio. Em muitos casos, trata-se apenas de uma pequena sombra. (Nem sempre alguém sai machucado.) Mentir para si mesmo, no recôndito dos próprios pensamentos, não obscurece a história da humanidade tanto quanto disseminar mentiras em público ou incitar violência. No entanto, uma parte de mim não consegue deixar de lamentar. E como não pretendo causar danos, mas apenas debater ideias, acredito ser correto e apropriado, como ser humano, me importar com meus semelhantes. Essa também é uma posição que pretendo defender no futuro.



**F - Política e Racionalidade**



## 57 - A política é o assassino da mente



É comum vermos pessoas perdendo a cabeça quando o assunto é política. As razões evolutivas para esse comportamento são claras e merecem análise: em tempos ancestrais, a política era questão de vida ou morte. Sexo, riqueza, aliados e reputação eram cruciais para a sobrevivência. Mesmo hoje, ao discutirmos temas como o aumento do salário mínimo, nosso cérebro usa as mesmas adaptações que nos ajudaram a sobreviver no passado. Estar do lado errado da discussão poderia significar a morte, enquanto estar do lado certo poderia ser a chave para eliminar nossos inimigos odiados!

Se você quer ressaltar a importância da ciência e da racionalidade, evite escolher tópicos políticos atuais, se possível. Se seu argumento se relaciona diretamente com a política, é melhor falar sobre Luís XVI durante a Revolução Francesa. A política é uma esfera vital para aplicarmos nossa racionalidade individual, mas é péssima para aprender ou debater racionalidade, a menos que todos os debatedores já sejam racionais. A política é a continuação da guerra por outros meios. Os argumentos são soldados.

Após decidir de que lado você está, é preciso apoiar todos os argumentos desse lado e atacar todos os do lado oposto; caso contrário, é como esfaquear seus próprios soldados pelas costas — auxiliando o inimigo. Pessoas normalmente equilibradas, que considerariam todos os lados de uma questão em suas vidas profissionais, como cientistas, podem de repente se transformar em zumbis repetindo slogans quando se trata de tomar partido pelo lado Azul ou Verde em uma questão política.

Na inteligência artificial, especialmente no raciocínio não monotônico, há um problema padrão: “Todos os Quakers são pacifistas. Nem todos os republicanos são pacifistas. Nixon é quaker e republicano. Nixon é um pacifista?”

Por que escolher esse exemplo? Para inflamar as emoções políticas dos leitores e distraí-los da questão principal? Para fazer com que republicanos se sintam indesejados em cursos de inteligência artificial e desencorajá-los a entrar no campo? (E não, não sou republicano. Nem democrata.)

Por que alguém usaria um exemplo tão provocativo para ilustrar o raciocínio não monotônico? Provavelmente porque o autor simplesmente não resistiu à tentação de criticar os odiados Verdes. É tão satisfatório dar uma boa alfinetada, você sabe, é como tentar resistir a um biscoito de chocolate.

Assim como os biscoitos de chocolate, nem tudo que dá prazer é bom para você.

Não estou afirmando que devemos ser apolíticos ou mesmo adotar o ideal de [ponto de vista neutro](#) da Wikipedia. Mas tente evitar críticas contundentes, se possível. Se o seu tópico está legitimamente relacionado à tentativa de proibir a evolução nos currículos escolares, vá em frente e fale sobre isso — mas não culpe explicitamente todo o Partido Republicano. Afinal, alguns de seus leitores podem ser republicanos e podem acreditar que o problema são apenas alguns indivíduos e não o partido inteiro. Da mesma forma que acontece com o princípio de ponto de vista neutro da Wikipedia, não importa se você acha que o Partido Republicano é realmente culpado. É melhor para o desenvolvimento da comunidade discutir o assunto sem evocar cores políticas.



## 58 - Debates sobre políticas não devem parecer unilaterais



Robin Hanson propôs [a criação de lojas onde produtos proibidos pudessem ser vendidos](#). Essa política é sustentada por diversos argumentos sólidos: o direito inerente à liberdade individual, o incentivo dos burocratas em proibir tudo e a tendência dos legisladores em serem tão parciais quanto os cidadãos. No entanto, mesmo assim (respondi), existe uma possibilidade preocupante: uma mãe pobre, honesta e com pouca educação poderia entrar nessas lojas e comprar um *Dr. Snakeoil's Sulfuric Acid Drink* (Elixir de ácido sulfúrico do Dr. Óleo de Cobra) para tratar sua artrite. O resultado trágico seria seu falecimento, deixando seus órfãos chorando na televisão nacional.

Minha intenção era apenas fazer uma observação factual simples. Então, por que algumas pessoas entenderam como um argumento a favor da regulamentação?

Em questões de fatos simples, como a origem da vida na Terra por seleção natural, é legítimo esperar que o argumento seja unilateral. Afinal, os próprios fatos vão por um caminho ou outro, e o chamado “equilíbrio de evidências” deve refletir isso. Na verdade, segundo a definição bayesiana de evidência, a “evidência forte” é apenas aquele tipo de evidência que esperamos encontrar em apenas um dos lados do argumento.

Contudo, não há motivo para ações complexas, com diversas consequências, exibirem essa propriedade unilateral. Então, por que as pessoas parecem desejar que seus debates políticos sejam unilaterais?

A resposta é simples: [a política é o assassino do pensamento](#). Os argumentos são soldados. Após escolher um lado, é preciso apoiar todos os argumentos desse lado e atacar todos os argumentos que pareçam favorecer o lado oposto. Fazer o contrário seria como esfaquear seus próprios soldados pelas costas. Se você seguir esse padrão, os debates políticos também parecerão unilaterais para você. Os custos e as desvantagens da sua política favorita tornam-se os soldados inimigos, a serem combatidos por todos os meios necessários.

É crucial estar ciente de um padrão falho relacionado: acreditar que a sabedoria profunda reside em comprometer-se com uma igualdade perfeita entre quaisquer duas posições políticas mais destacadas. A realidade é que uma política pode legitimamente ter custos ou benefícios desiguais. Se as questões políticas não se inclinarem para um lado ou para o outro, seria impossível tomar decisões a respeito delas. No entanto, há uma tendência humana de negar todos os custos de uma política favorecida ou negar todos os benefícios de uma política desfavorecida. Como resultado, as pessoas tendem a pensar que as compensações políticas são muito mais desequilibradas do que realmente são.

Consideremos o exemplo inicial: se permitirmos a existência de lojas que vendem produtos proibidos, é provável que alguma mãe pobre, honesta e com pouca instrução, com cinco filhos, adquira algo que a leve à morte. Essa é uma previsão baseada em uma consequência factual e, como tal, parece bastante direta. Uma pessoa racional prontamente concordaria que isso é verdade, independentemente da postura que ela adote em relação à questão política. Podemos considerar outros fatores: tornar as coisas ilegais apenas as encarece, os reguladores podem abusar de seu poder, ou a liberdade individual dela prevalece sobre o desejo de interferir em sua vida. Mas a dura realidade permanece: ela ainda morrerá.

Enfrentamos uma verdade incômoda: vivemos em um universo injusto. Como todos os primatas, os seres humanos têm fortes reações negativas diante da percepção de injustiça. Consequentemente, consideramos esse fato estressante. Existem dois métodos populares para lidar com a dissonância cognitiva resultante. Primeiramente, pode-se alterar a percepção dos fatos — negando que eventos injustos tenham ocorrido ou editando a história para fazê-la parecer justa. (Isso é mediado pela [heurística do afeto](#) e pela [falácia do mundo](#)

[justo.](#)) Em segundo lugar, pode-se mudar a própria moralidade — negando que os eventos sejam injustos.

Alguns libertários argumentam: se alguém entrar em uma “loja de produtos proibidos”, passando por rótulos de advertência claros que afirmam que os itens NESSA loja podem ser letais, e comprar algo que cause a própria morte, então a culpa é exclusivamente da pessoa e ela merece as consequências. Se isso fosse uma verdade moral, não haveria desvantagens em ter lojas que vendem produtos proibidos. Seria apenas um benefício líquido; uma troca unilateral sem inconvenientes.

Por outro lado, alguns argumentam que os reguladores podem ser treinados para fazer escolhas racionais e conforme os interesses dos consumidores. Se esses fossem os fatos em questão, então não haveria desvantagem moral na regulamentação.

Goste ou não, existe uma loteria do nascimento quando se trata de inteligência - embora esse seja um dos casos em que a injustiça do universo é tão extrema que muitas pessoas preferem negar os fatos. A evidência experimental é esmagadora: há um componente puramente genético de 0,6 a 0,8. Mas mesmo que isso seja negado, não se pode escolher a educação fornecida pelos pais ou a escola primária que frequentamos.

Minha criação me ensinou que negar a realidade é um erro moral. Se eu me envolvesse em um otimismo irracional de que beber ácido sulfúrico provavelmente me traria benefícios, estaria indo contra o que fui advertido e considerado inaceitável. Algumas pessoas nascem em ambientes — não vamos discutir seus genes, pois essa parte é extremamente injusta — onde o líder local lhes diz que ter fé é correto e ser cético é errado. De boa vontade, elas seguem esse conselho e acabam morrendo. Ao contrário de você, elas não foram criadas acreditando que as pessoas são responsáveis por suas escolhas individuais de seguir o exemplo da sociedade. Você realmente acredita que é tão perspicaz que teria sido um cético científico adequado mesmo se tivesse nascido no ano 500 EC? Sim, existe uma loteria do nascimento, independentemente de suas crenças em relação aos genes.

Afirmar “Pessoas que comprem produtos perigosos merecem se machucar!” não é ser obstinado. É uma forma de se recusar a viver em um universo injusto. A verdadeira obstinação está em dizer: “Sim, o ácido sulfúrico causa uma morte horrível e dolorosa, e não, aquela mãe de cinco filhos não merecia isso, mas vamos manter as lojas abertas de qualquer maneira porque fizemos esse cálculo de custo-benefício.” Você consegue imaginar um político dizendo isso? Nem eu. No entanto, enquanto os economistas têm influência sobre as políticas, pode ser útil que eles pensem nisso em particular — talvez até mesmo expressando essa ideia em artigos de jornais, habilmente envolvidos em linguagem polissilábica para evitar que a mídia os cite diretamente.

Eu não acredito que a morte de alguém como resultado de uma escolha estúpida seja motivo de celebração. Vejo isso como uma tragédia. Nem sempre ajudamos as pessoas protegendo-as das consequências de suas próprias ações, mas traço uma linha moral quando se trata da pena de morte. Se você está morto, não pode aprender com seus erros.

Infelizmente, o universo não compartilha da minha opinião. Veremos qual de nós ainda estaremos de pé quando tudo isso acabar.

## 59 - A balança da Justiça, o caderno da Racionalidade



[Lady Justice](#) (Senhora Justiça) é frequentemente representada segurando uma balança. Um conjunto de balanças possui a propriedade de que tudo o que puxa um lado para baixo empurra o outro lado para cima. Isso torna as coisas muito convenientes e fáceis de acompanhar. No entanto, essa representação é geralmente uma distorção grosseira da realidade.

No discurso humano, há uma tendência natural de tratar a discussão como uma forma de combate, uma extensão da guerra, um esporte. Nos esportes, basta acompanhar quantos pontos cada equipe marcou. Há apenas dois lados, e cada ponto marcado contra um lado é um ponto a favor do outro. Todos na plateia mantêm uma contagem mental de quantos pontos cada orador marca contra o outro. No final do debate, o orador que acumulou mais pontos é considerado o vencedor. Portanto, tudo o que o orador vencedor diz deve ser verdadeiro, e tudo o que o perdedor diz deve estar errado.

Um estudo intitulado *The Affect Heuristic in Judgments of Risks and Benefits* (A heurística do afeto no julgamento de riscos e benefícios) investigou se os participantes misturavam seus julgamentos sobre os possíveis benefícios de uma determinada tecnologia (como a energia nuclear) e os possíveis riscos associados a essa tecnologia, resultando em um sentimento geral positivo ou negativo em relação à tecnologia [1]. Suponha que eu lhe diga inicialmente que um determinado tipo de reator nuclear gera menos resíduos nucleares do que projetos concorrentes. No entanto, em seguida, eu menciono que esse reator é mais instável do que os projetos concorrentes, com um maior risco de fusão caso várias coisas saiam errado simultaneamente.

Se o reator tiver uma probabilidade maior de derreter, isso pode parecer um “ponto negativo” para o reator ou para aqueles que defendem sua construção. E se o reator produzir menos resíduos, isso pode ser considerado um “ponto positivo” para o reator ou para a sua construção. Então, esses dois fatos são opostos entre si? Não, no mundo real não são. Embora esses dois fatos possam ser citados por diferentes lados do mesmo debate, eles são logicamente distintos; os fatos não têm preferência.

Se é um fato físico sobre o projeto do reator que ele é passivamente seguro (não se tornará supercrítico mesmo se ocorrerem falhas nos sistemas de refrigeração, por exemplo), isso não implica necessariamente que o reator gerará menos resíduos ou produzirá eletricidade a um custo mais baixo. Todas essas características seriam benéficas, mas não são o mesmo. A quantidade de resíduos gerados pelo reator é determinada pelas suas próprias propriedades, enquanto outras características físicas do reator podem tornar a reação nuclear mais instável. Mesmo que algumas das mesmas características do projeto estejam envolvidas, é necessário considerar separadamente a probabilidade de colapso e o volume esperado de resíduos gerados. São duas questões físicas distintas, com respostas factuais diferentes.

Mas estudos como o mencionado mostram que as pessoas tendem a avaliar tecnologias — e muitos outros problemas — com base em uma impressão geral positiva ou negativa. Se informarmos às pessoas que um determinado projeto de reator gera menos resíduos, elas tendem a considerar a probabilidade de derretimento como sendo menor. Isso resulta em uma resposta incorreta para questões físicas com respostas factuais definidas, porque mistura questões logicamente distintas — tratando fatos como soldados em lados opostos de uma guerra, e acreditando que um soldado de um lado pode lutar contra qualquer soldado do outro lado.

Um conjunto de balanças não é totalmente inadequado para a representação de *Lady Justice* quando ela está investigando uma questão estritamente factual de culpa ou inocência. Por exemplo, se Fulano matou

Sicrano ou não. Conforme ensinado por E. T. Jaynes, toda evidência bayesiana consiste em fluxos de probabilidade entre diferentes hipóteses. Não existe evidência que “apoie” ou “contradiga” uma única hipótese, a menos que outras hipóteses se saiam pior, ou melhor. Portanto, se *Lady Justice* está investigando uma única questão estritamente factual com uma resposta binária, um conjunto de balanças seria uma ferramenta apropriada. No entanto, se a *Lady Justice* precisa considerar questões mais complexas, ela deve renunciar às balanças ou à espada.

Nem todos os argumentos se reduzem a extremos opostos. *Lady Racionalidade* carrega consigo um caderno, onde ela registra todos os fatos que não estão a favor de nenhum lado em particular.

## Referências

[1] Melissa L. Finucane et al., “The Affect Heuristic in Judgments of Risks and Benefits,” *Journal of Behavioral Decision Making* 13, no. 1 (2000): 1–17.

## 60 - Viés de correspondência



O viés de correspondência é a tendência de tirar conclusões sobre as características pessoais únicas e duradouras de alguém com base em comportamentos que podem ser plenamente explicados pelas circunstâncias em que ocorrem<sup>13</sup>.

— Gilbert e Malone [1]

Tendemos a fazer uma correspondência direta entre as ações e personalidades de outras pessoas. Quando vemos alguém chutar uma máquina de venda automática sem motivo aparente, presumimos que essa pessoa tem um temperamento irritado. No entanto, quando nós mesmos chutamos a máquina, é porque o ônibus estava atrasado, o trem chegou cedo, nosso relatório estava atrasado e a maldita máquina comeu nosso dinheiro do almoço pelo segundo dia consecutivo. Certamente, pensamos, qualquer um chutaria a máquina de venda automática nessa situação.

Atribuimos nossas próprias ações às circunstâncias, vendo nossos comportamentos como respostas perfeitamente normais às experiências que enfrentamos. No entanto, quando outra pessoa chuta uma máquina de venda automática, não conseguimos ver sua história pessoal por trás disso. Só vemos o chute, sem um motivo aparente, e presumimos que essa pessoa é naturalmente irritada, já que agiu sem provocação.

Devemos considerar as probabilidades. Há mais atrasos de ônibus no mundo do que pessoas com mutações genéticas que as levam a chutar máquinas de venda automática espontaneamente devido a altos níveis de raiva. Em média, cada indivíduo possui de duas a dez mutações genéticas, embora seja improvável que qualquer uma delas afete diretamente o comportamento. Da mesma forma, os traços de personalidade de alguém provavelmente não estão muito longe da média. Sugerir o oposto seria assumir uma improbabilidade desnecessária.

Mesmo quando as pessoas são explicitamente informadas sobre as causas situacionais, elas parecem não levar adequadamente em consideração o comportamento observado. Por exemplo, quando os participantes são informados de que um orador foi designado aleatoriamente para falar sobre uma posição pró-aborto ou antiaborto, eles ainda acreditam que os oradores têm tendências alinhadas com a posição designada, mesmo que tenha sido uma atribuição aleatória. [2]

Parece bastante intuitivo explicar a chuva pelos espíritos da água; explicar o fogo por meio de um material inflamável (flogisto) escapando da matéria em combustão; explicar o efeito sedativo de um medicamento dizendo que ele contém uma “propriedade indutora do sono”. A realidade geralmente envolve mecanismos mais complexos: um ciclo de evaporação e condensação subjacente à chuva, uma reação oxidante subjacente ao fogo e interações químicas com o sistema nervoso para sedativos. No entanto, esses mecanismos parecem mais complexos do que as essências; eles são mais difíceis de compreender e menos acessíveis. Portanto, quando alguém chuta uma máquina de venda automática, presumimos que essa pessoa tem uma tendência inata de chutar máquinas de venda automática.

A menos que sejamos nós mesmos a pessoa que chuta a máquina — nesse caso, estamos agindo perfeitamente normalmente, dadas nossas circunstâncias; certamente qualquer outra pessoa faria o mesmo. Na verdade, superestimamos a probabilidade de os outros responderem da mesma maneira que nós — o

---

13 NT: Texto original em inglês. *The correspondence bias is the tendency to draw inferences about a person's unique and enduring dispositions from behaviors that can be entirely explained by the situations in which they occur.*

chamado “efeito de consenso ilusório”. Estudantes que bebem tendem a superestimar consideravelmente a proporção de colegas que também bebem, enquanto os que não bebem subestimam consideravelmente essa proporção. O “erro de atribuição fundamental” refere-se à nossa tendência de atribuir o comportamento dos outros às suas características pessoais, enquanto invertemos essa tendência quando se trata de nós mesmos.

Para entender por que as pessoas agem da maneira que agem, devemos primeiro perceber que todos se veem como se estivessem agindo normalmente. Não devemos questionar quais características estranhas e mutáveis elas possuem desde o nascimento, pois isso corresponde diretamente ao seu comportamento superficial. Em vez disso, devemos perguntar em quais situações as pessoas se encontram. Sim, as pessoas têm características pessoais, mas não existem peculiaridades hereditárias suficientes para explicar diretamente todos os comportamentos superficiais que observamos.

Suponha que eu lhe dê um controle com dois botões, um vermelho e um verde. O botão vermelho destrói o mundo e o botão verde impede que o botão vermelho seja pressionado. Qual botão você apertaria? O verde. Qualquer pessoa que dê uma resposta diferente provavelmente está [tornando a questão excessivamente complicada](#).

No entanto, às vezes as pessoas me questionam sobre o motivo de eu querer [salvar o mundo](#). Como se eu tivesse passado por uma infância traumática ou algo assim. Na verdade, parece uma decisão bastante óbvia... se você enxergar a situação por essa perspectiva.

Posso ter visões que fogem do comum e exigem uma explicação: por que acredito nessas coisas quando a maioria das pessoas não acredita? No entanto, considerando essas crenças, minha reação não parece requerer uma explicação extraordinária. Talvez eu esteja sujeito a um falso consenso; talvez eu esteja superestimando quantas pessoas pressionariam o botão verde se vissem a situação nesses termos. Mas, sabe, eu ainda apostaria que haveria pelo menos uma minoria significativa.

A maioria das pessoas se percebe como perfeitamente normais por dentro. Até mesmo aquelas pessoas que você odeia, as que cometem atos terríveis, não são mutantes excepcionais. Nenhuma mutação é necessária, infelizmente. Quando você compreender isso, estará pronto para deixar de se surpreender com os eventos humanos.

## Referências

- [1] Daniel T. Gilbert and Patrick S. Malone, “The Correspondence Bias,” *Psychological Bulletin* 117, no. 1 (1995): 21–38, [http://www.wjh.harvard.edu/~dtg/Gilbert%20&%20Malone20\(CORRESPONDENCE%20BIAS\).pdf](http://www.wjh.harvard.edu/~dtg/Gilbert%20&%20Malone20(CORRESPONDENCE%20BIAS).pdf).
- [2] Edward E. Jones and Victor A. Harris, “The Attribution of Attitudes,” *Journal of Experimental Social Psychology* 3(1967):1–24, [http://www.radford.edu/~jaspelme/443/spring-2007/Articles/Jones\\_n\\_Harris\\_1967.pdf](http://www.radford.edu/~jaspelme/443/spring-2007/Articles/Jones_n_Harris_1967.pdf).

## 61 - Seus inimigos são naturalmente maus?



Observamos uma conexão muito direta entre as ações dos outros e suas disposições inatas. Em vez de questionar situações reais ou imaginárias que poderiam explicar o comportamento, enxergamos disposições incomuns que correspondem exatamente ao comportamento incomum. Levantamos a hipótese da existência de mutantes.

Quando alguém nos ofende verdadeiramente — realizando uma ação que desaprovamos (com ou sem razão) —, percebo que o viés de correspondência se intensifica. Há uma forte tendência em atribuir ações negativas ao caráter mutante e maligno do *Inimigo*. Não se trata de um ponto moral, mas de uma questão estritamente relacionada à probabilidade anterior. Devemos nos perguntar o que o Inimigo pode acreditar sobre sua situação, o que reduziria a aparente estranheza de seu comportamento. Isso nos permitiria supor uma disposição menos excepcional e, assim, lidar com uma carga menor de improbabilidade.

Em 11 de setembro de 2001, dezenove homens muçulmanos sequestraram quatro aviões a jato em um esforço deliberadamente suicida para prejudicar os Estados Unidos da América. Agora, por que você acha que eles poderiam ter feito isso? Porque viam os EUA como um farol de liberdade para o mundo, mas nasceram com uma disposição mutante que os fazia odiar a liberdade?

Realisticamente, a maioria das pessoas não constrói suas histórias de vida colocando-se como os vilões. Todos são heróis em suas próprias histórias. A história do Inimigo, vista pelo próprio Inimigo, não o retratará negativamente. Se tentarmos interpretar as motivações que tornariam o *Inimigo* uma figura má, acabaremos enganando-nos sobre o que realmente se passa em sua mente.

Mas [a política é a assassina da mente](#). O debate é uma guerra; os argumentos são soldados. Uma vez que você tenha escolhido de que lado está, deve apoiar todos os argumentos desse lado e atacar todos os argumentos que parecem favorecer o lado oposto. Caso contrário, é como esfaquear seus próprios soldados pelas costas.

Se o *Inimigo* tiver uma disposição maligna, isso seria um argumento a favor do seu lado. E qualquer argumento que beneficie seu lado deve ser apoiado, independentemente de quão tolo seja — caso contrário, você estará aliviando a pressão em algum ponto da frente de batalha. Todos se esforçam para superar o outro na denúncia patriótica e ninguém se atreve a contradizer. Logo, o *Inimigo* é retratado com chifres, asas de morcego, hálito flamejante e presas que pingam veneno corrosivo. Se você negar qualquer aspecto disso com base apenas em fatos, estará defendendo o lado do Inimigo; você será um traidor. Pouquíssimas pessoas entenderão que você não está defendendo o Inimigo, apenas defendendo a verdade.

Se fosse necessário ser um mutante para realizar ações monstruosas, a história da humanidade seria muito diferente. Os mutantes seriam raros.

Ou talvez haja o medo de que compreender leve ao perdão. É mais fácil eliminar mutantes malignos. É mais inspirador gritar em guerra: “Morra, escória viciosa!” em vez de “Morram, pessoas que poderiam ter sido como eu, mas cresceram em um ambiente diferente!” Você pode se sentir culpado por matar pessoas que não eram puramente do mal.

Isso me parece um anseio profundo por um [debate político unilateral](#), onde a melhor política não tem desvantagens. Se um exército está atravessando a fronteira ou um lunático está avançando em sua direção com uma faca, as opções políticas são (a) se defender ou (b) se render e morrer. Se você se defender, pode ter

que matar. Se você matar alguém que, em outro mundo, poderia ter sido seu amigo, isso é uma tragédia. E é uma tragédia. A outra opção, se render e morrer, também é uma tragédia. Por que deveria existir uma opção não trágica? Quem disse que a melhor estratégia disponível não deve ter desvantagens? Se alguém tem que morrer, então pode muito bem ser o iniciador da violência, para desencorajar futuros atos de violência e, assim, minimizar o total de mortes.

Se o *Inimigo* possui uma disposição comum e está agindo com base em crenças sobre sua situação que tornariam a violência uma resposta típica humana, isso não significa que suas crenças sejam factualmente precisas. Isso não significa que eles estejam justificados. Significa apenas que você terá que confrontar alguém que é o herói de sua própria história, onde o protagonista morrerá na página 80. Isso é uma tragédia, mas é melhor do que uma tragédia alternativa. É a escolha que todo policial faz todos os dias, para evitar que nossos pequenos mundos organizados se dissolvam no caos.

Quando você tem uma compreensão precisa da psicologia do Inimigo — quando você sabe o que realmente se passa em sua mente — esse conhecimento não lhe proporcionará uma satisfação deliciosa em atingir o lado oposto. Não lhe dará uma sensação calorosa de indignação justa. Não fará você se sentir bem consigo mesmo. Se sua compreensão o deixa insuportavelmente triste, pode ser que você esteja vendo o mundo como ele realmente é. Mais raramente, uma compreensão precisa pode causar arrepios de horror, como ao lidar com verdadeiros psicopatas ou pessoas neurologicamente intactas com crenças que destruíram completamente sua sanidade (Cientólogos ou *Jesus Campers*<sup>14</sup>).

Então, vamos direto ao ponto: os sequestradores do 11 de setembro não eram mutantes malignos. Eles não odiavam a liberdade. Eles também eram os heróis de suas próprias histórias e morreram pelo que acreditavam ser certo — verdade, justiça e o caminho islâmico. Se os sequestradores se viam dessa forma, isso não significa que suas crenças eram verdadeiras. Se os sequestradores se viam dessa forma, isso não significa que temos que concordar que o que eles fizeram foi justificado. Se os sequestradores se viram dessa forma, isso não significa que os passageiros do voo 93 da *United*<sup>15</sup> deveriam ter se rendido e deixado que isso acontecesse. Isso significa apenas que em outro mundo, se eles tivessem sido criados em um ambiente diferente, esses sequestradores poderiam ter se tornado policiais. E isso, de fato, é uma tragédia. Bem-vindo à Terra.

---

14 NT: Cristãos fundamentalistas.

15 NT: Companhia aérea norte-americana *United Airlines*.



## 62 - A estupidez ao contrário não é inteligência



“... Então nossa equipe nesta linha do tempo começou a trabalhar em ação corretiva. Aqui.”

Ele limpou a tela e começou a digitar freneticamente. Página após página surgiam, trazendo relatos de pessoas que afirmavam ter visto os misteriosos discos voadores, e cada relato era mais fantástico que o anterior.

“A técnica padrão de desacreditar”, Verkan Vall sorriu. “Eu ouvi falar um pouco sobre os ‘discos voadores’, e tudo isso era apenas uma brincadeira. Nesse tipo de cultura, você sempre pode desacreditar uma história verdadeira ao estabelecer dez outras, claramente falsas, paralelas a ela.”<sup>16</sup>

— H. Beam Piper, *Police Operation* (Operação policial) [1]

Piper estava certo. Pessoalmente, não acredito que haja alienígenas escondidos infiltrados por aqui. Mas minha descrença não tem nada a ver com a terrível e constrangedora irracionalidade das seitas aos discos voadores — pelo menos, eu espero que não.

Você e eu acreditamos que as seitas que cultuam discos voadores surgiram mesmo na completa ausência de quaisquer discos voadores. [As seitas podem surgir em torno de praticamente qualquer ideia](#), graças à tolice humana. Essa bobagem opera em uma direção diferente da intervenção alienígena: esperaríamos ver seitas de discos voadores, independentemente da existência real dos mesmos. Mesmo que houvesse alienígenas escondidos, não seria menos provável que surgissem seitas de discos voadores. A probabilidade condicional P (seitas | alienígenas) não é menor que P (seitas | ~alienígenas), a menos que se suponha que os alienígenas escondidos estejam deliberadamente suprimindo as seitas de discos voadores. Conforme a definição bayesiana de evidência, a observação de “existem seitas de discos voadores” não é evidência contra a existência de discos voadores. Não há muitas evidências em uma direção ou outra.

Isso é uma aplicação do princípio geral expresso por Robert Pirsig, que diz: “O maior tolo do mundo pode dizer que o Sol está brilhando, mas isso não o torna escuro.”<sup>17</sup> [2]

Se você conhece alguém que acerta apenas 0,01% das vezes em questões de sim ou não, você pode obter 99,99% de precisão apenas invertendo suas respostas. Eles precisariam fazer todo o trabalho de obter boas evidências enraizadas na realidade e processar essas evidências de maneira coerente, apenas para confiavelmente obter resultados opostos. Eles teriam que ser superinteligentes para serem tão estúpidos.

Um carro com o motor quebrado não pode retroceder a 200 mph (aprox. 322 km/h), mesmo que o motor esteja realmente quebrado.

Se a estupidez não se correlaciona consistentemente com a verdade, muito menos a maldade humana deveria se correlacionar com a verdade. O oposto do [efeito halo \(auréolas\)](#) é o efeito chifres: todas

---

16 NT: Texto original em inglês. “... Then our people on that time-line went to work with corrective action. Here.” *He wiped the screen and then began punching combinations. Page after page appeared, bearing accounts of people who had claimed to have seen the mysterious disks, and each report was more fantastic than the last. “The standard smother-out technique,” Verkan Vall grinned. “I only heard a little talk about the ‘flying saucers,’ and all of that was in joke. In that order of culture, you can always discredit one true story by setting up ten others, palpably false, parallel to it.”*

17 NT: Texto original em inglês. “The world’s greatest fool may say the Sun is shining, but that doesn’t make it dark out.”

as qualidades negativas percebidas estão correlacionadas. Se Stalin é mal, então tudo o que ele diz deve ser falso. Você não gostaria de concordar com Stalin, não é?

Stalin também acreditava que  $2 + 2 = 4$ . No entanto, se você defender qualquer afirmação feita por Stalin, inclusive “ $2 + 2 = 4$ ”, as pessoas apenas verão que você está “concordando com Stalin”; você será colocado ao lado dele.

Corolários desse princípio:

- Para argumentar honestamente contra uma ideia, você deve contestar os melhores argumentos dos defensores mais fortes. Argumentar contra defensores mais fracos não prova nada, pois até mesmo a ideia mais forte atrairá defensores fracos. Se você deseja argumentar contra o transumanismo ou a singularidade da inteligência, você deve desafiar diretamente os argumentos de Nick Bostrom ou Eliezer Yudkowsky pós-2003. [O caminho menos conveniente](#) é o único válido.
- Apresentar lunáticos tristes e patéticos, levados à loucura por sua interpretação de uma ideia, não é evidência contra essa ideia. Muitos seguidores da Nova Era enlouqueceram ainda mais devido à sua própria interpretação pessoal da [mecânica quântica](#).
- Alguém disse uma vez: “Nem todos os conservadores são estúpidos, mas a maioria das pessoas estúpidas são conservadoras”. Se você não consegue adotar uma mentalidade em que essa afirmação, verdadeira ou falsa, pareça completamente irrelevante como crítica ao conservadorismo, então você não está preparado para pensar racionalmente sobre política.
- O argumento *ad hominem* é inválido.
- Você precisa conseguir argumentar contra o genocídio sem mencionar “Hitler queria exterminar os judeus”. Se Hitler não tivesse defendido o genocídio, ainda seria errado, certo?
- Em termos hansonianos: sua disposição instintiva de acreditar em algo mudará com sua disposição de se associar a pessoas conhecidas por acreditarem nisso, independentemente de a crença ser verdadeira ou não. Algumas pessoas podem relutar em acreditar que Deus não existe, não porque existam evidências de que Deus existe, mas porque hesitam em se alinhar com Richard Dawkins ou com aqueles “ateus estridentes” que publicamente afirmam “Deus não existe”.
- Se o seu computador atual parar de funcionar, você não pode concluir que tudo no sistema atual está errado e que você precisa de um novo sistema sem processador AMD, placa de vídeo ATI, disco rígido Maxtor ou ventoinhas, mesmo que o seu sistema atual tenha todos esses componentes e não funcione. Talvez você apenas precise de um novo cabo de alimentação.
- Se cem inventores falharem em construir máquinas voadoras usando metal, madeira e lona, isso não significa que você precise realmente de uma máquina voadora feita de carne e osso. Se mil projetos falharem em construir Inteligência Artificial usando computação baseada em eletricidade, isso não significa que a eletricidade seja a fonte do problema. Até que você entenda o problema, é altamente improvável que soluções esperançosas atinjam a resolução adequada.

## Referências

[1] Henry Beam Piper, “Police Operation,” *Astounding Science Fiction* (July 1948).

[2] Robert M. Pirsig, *Zen and the Art of Motorcycle Maintenance: An Inquiry Into Values*, 1st ed. (New York: Morrow, 1974).

## 63 - Argumento bloqueia autoridade



Cenário 1: Barry é um geólogo renomado. Charles é um jovem delinquente de quatorze anos com um longo histórico de prisões e ocasionais episódios psicóticos. Barry faz uma declaração categoricamente contraintuitiva sobre rochas a Arthur, que acredita ser 90% provável. Em seguida, Charles faz uma declaração igualmente contraintuitiva sobre rochas, e Arthur acredita que é apenas 10% provável. É claro que Arthur está considerando a autoridade do falante para decidir se deve ou não acreditar em suas afirmações.

Cenário 2: David faz uma declaração pouco intuitiva sobre física e oferece a Arthur uma explicação detalhada de seus argumentos, incluindo referências. Por outro lado, Ernie também faz uma afirmação pouco intuitiva, porém apresenta um argumento pouco convincente com diversos saltos de fé. Tanto David quanto Ernie afirmam que estão dando a melhor explicação possível (não apenas para Arthur, mas para qualquer pessoa). Após ouvir a explicação de David, Arthur atribui uma probabilidade de 90% à sua afirmação, enquanto atribui apenas 10% à declaração de Ernie.

À primeira vista, esses dois cenários podem parecer simétricos: ambos envolvem considerar evidências úteis, seja em relação a uma autoridade forte ou fraca, ou um argumento forte, ou fraco.

Agora, suponha que Arthur peça a Barry e Charles para apresentarem casos técnicos completos com referências, e que ambos apresentem casos igualmente bons. Em seguida, Arthur verifica as referências e confirma que estão corretas. Depois, Arthur pede a David e Ernie para informarem suas credenciais e descobre que ambos têm credenciais semelhantes — talvez ambos sejam palhaços ou ambos sejam físicos.

Assumindo que Arthur tenha conhecimento suficiente para compreender todos os argumentos técnicos — caso contrário, eles não passam de ruídos impressionantes —, ele deve reconhecer que David tem uma grande vantagem em plausibilidade em relação à Ernie, enquanto Barry, na melhor das hipóteses, tem uma pequena vantagem sobre Charles.

Na verdade, se os argumentos técnicos forem suficientemente bons, a vantagem de Barry sobre Charles pode não valer a pena ser considerada. Um argumento técnico de qualidade é aquele que reduz a confiança na autoridade pessoal do orador.

Da mesma forma, se acreditarmos que Ernie deu o melhor argumento possível, considerando todas as etapas inferenciais que ele executou e todo o suporte que ele considerou — citando quaisquer autoridades que ele possa ter consultado — podemos praticamente ignorar qualquer informação sobre suas credenciais. Independentemente de Ernie ser um físico ou um palhaço, isso não importa. É claro que essa suposição só é válida se tivermos conhecimento técnico suficiente para avaliar o argumento. Caso contrário, Ernie estaria simplesmente pronunciando sílabas místicas, e se acreditamos nessas sílabas dependeria muito de sua autoridade.

Parece haver uma assimetria entre argumento e autoridade. Se conhecermos a autoridade, ainda teremos interesse em ouvir seus argumentos. Por outro lado, se conhecermos completamente os argumentos, teremos pouco a aprender com a autoridade.

É evidente (diz o novato) que autoridade e argumento são tipos de evidência fundamentalmente diferentes, uma discrepância inexplicável para os métodos rigorosamente limpos da teoria de probabilidade bayesiana. Embora a força da evidência — 90% contra 10% — seja igual em ambos os casos, elas não se comportam de maneira semelhante quando combinadas. Como vamos contabilizar isso?

Aqui está uma prova técnica de como expressar essa diferença na teoria da probabilidade. (O resto pode ser baseado em minha autoridade pessoal ou em referências de consulta.)

Se  $P(H|E1) = 90\%$  [A probabilidade de H dado E1 é igual a noventa por cento] e  $P(H|E2) = 9\%$  [a probabilidade de H dado E2 é igual a nove por cento], qual é a probabilidade  $P(H|E1,E2)$  [de H dado E1 e E2]?

Se saber que E1 é verdadeiro nos leva a atribuir 90% de probabilidade a H, e sabermos que E2 é verdadeiro nos leva a atribuir 9% de probabilidade a H, então qual probabilidade devemos atribuir a H se descobrirmos tanto E1 quanto E2 juntos?

Isso não pode ser calculado na teoria da probabilidade apenas com as informações fornecidas. A informação que está faltando não é a probabilidade anterior de H. Os eventos E1 e E2 podem ser dependentes um do outro.

Suponhamos que H seja “A minha calçada está escorregadia”, E1 seja “O meu aspersor está funcionando” e E2 seja “É noite”. A minha calçada fica escorregadia a partir de um minuto após o início do aspersor, até logo após o término do aspersor, e o aspersor funciona por dez minutos. Então, se sabemos que o aspersor está ligado, a probabilidade de a calçada estar escorregadia é de 90%. O aspersor está ligado durante 10% da noite, então se soubermos que é noite, a probabilidade de a calçada estar escorregadia é de 9%. Se soubermos que é noite e o aspersor está ligado, ou seja, se soubermos os dois fatos, a probabilidade de a calçada estar escorregadia é de 90%.

Podemos representar isso em um modelo gráfico da seguinte forma:

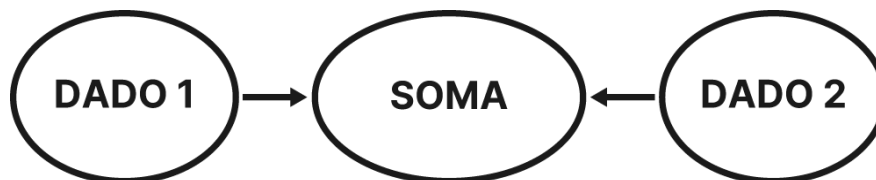
Se é noite ou não, determina se o aspersor está ligado ou desligado. E, se o aspersor estiver ligado, isso influencia se a calçada fica escorregadia ou não.

A direção das setas é significativa. Digamos que tivemos:

Isso significa que, se eu não tivesse conhecimento prévio sobre o aspersor, a probabilidade de que seja noite e de que a calçada esteja escorregadia seriam independentes entre si. Por exemplo, imagine que eu role dois dados e some os números obtidos para obter uma soma:



Se eu não informar a soma dos dois números e apenas disser que o primeiro dado marcou 6, isso não me diz nada sobre o resultado do segundo dado. No entanto, se eu disser que a soma dos dois dados é 7, então é possível concluir que o segundo dado deve ter exibido o número 1.



Determinar a dependência ou independência entre diferentes informações, dada uma série de conhecimentos prévios, é um tópico bastante técnico. Para aprofundar o conhecimento sobre o assunto, recomenda-se a leitura de dois livros específicos: [Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference \[1\]](#) (Raciocínio Probabilístico em Sistemas Inteligentes: Redes de Inferência Plausível) e [Causality: Models, Reasoning, and Inference](#) (Causalidade: Modelos, Raciocínio e Inferência), de Judea Pearl. [2] (Caso você não tenha tempo, é recomendável priorizar a leitura do primeiro livro.)

Se você tem conhecimento em leitura de diagramas causais, ao observar o diagrama das jogadas de dados, é possível identificar imediatamente que:

$$P(\text{Dado 1, Dado 2}) = P(\text{Dado 1}) \times P(\text{Dado 2})$$

$$P(\text{Dado 1, Dado 2} \mid \text{Soma}) \neq P(\text{Dado 1} \mid \text{Soma}) \times P(\text{Dado 2} \mid \text{Soma})$$

Ao observar o diagrama correto da calçada, é possível identificar fatos como:

$$P(\text{Escorregadio} \mid \text{Noite}) \neq P(\text{Escorregadio})$$

$$P(\text{Escorregadio} \mid \text{Aspersor}) \neq P(\text{Escorregadio})$$

$$P(\text{Escorregadio} \mid \text{Noite, Aspersor}) = P(\text{Escorregadio} \mid \text{Aspersor})$$

Em outras palavras, a probabilidade da calçada estar escorregadia, dada a informação sobre o aspersor e a noite, é a mesma probabilidade que atribuiríamos se tivéssemos conhecimento apenas sobre o aspersor. Isso significa que, ao incluir a informação do aspersor, a informação sobre a noite se tornou irrelevante para as inferências sobre a escorregabilidade da calçada.

Esse fenômeno é conhecido como *screening off*, ou *bloqueio* em português, e o critério que nos permite identificar essas independências condicionais a partir de diagramas causais é chamado de *D-separation* ou *D-separação*.

No caso de argumento e autoridade, o diagrama causal se parece com isso:



Se algo é verdadeiro, é provável existirem argumentos a favor, e os especialistas tendem a avaliar essas evidências e mudar suas opiniões (pelo menos em teoria!).

Se observarmos que um especialista acredita em algo, podemos inferir que existem evidências subjacentes (mesmo que não saibamos exatamente quais são essas evidências), e a partir da existência dessas evidências, podemos inferir a verdade da proposição.

No entanto, se soubermos o valor do nodo “Qualidade do argumento”, isso “D-separa” o nodo “Verdade” do nodo “Crença do Especialista”, bloqueando todas as rotas entre eles. Esse bloqueio segue certos critérios técnicos que, neste caso, parecem bastante evidentes.

Assim, mesmo, mesmo sem verificar a distribuição de probabilidade exata, podemos inferir diretamente do diagrama que:

$$P(\text{verdade} \mid \text{argumento, especialista}) = P(\text{verdade} \mid \text{argumento})$$

Essa não é uma contradição da teoria da probabilidade comum, mas sim uma forma mais simplificada de expressar fatos probabilísticos. As mesmas igualdades e desigualdades podem ser lidas em uma distribuição de probabilidade mais complexa, mas seria mais difícil de identificar visualmente. Assim como os aspersores e a luz do sol não são feitos de materiais ontologicamente diferentes, a autoridade e o argumento não exigem dois tipos diferentes de probabilidade.

Na prática, nunca é possível *eliminar completamente* a confiança na autoridade. As autoridades con-

fiáveis têm maior probabilidade de conhecer evidências contrárias que devem ser consideradas; enquanto autoridades menos confiáveis têm menos probabilidade de saber disso, tornando seus argumentos menos confiáveis. Não é possível eliminar esse fator apenas ouvindo as evidências que *eles* consideraram.

Transformar argumentos em equações matemáticas pode ser uma tarefa muito difícil. Afinal, avaliar a força de uma inferência pode depender de intuições que você não pode replicar sem a mesma experiência de trinta anos.

É inevitável que atribuamos uma probabilidade *ligeiramente* maior ao que E.T. Jaynes diz sobre probabilidade bayesiana do que a Eliezer Yudkowsky, mesmo que ambos afirmem o mesmo. Os cinquenta anos adicionais de experiência de Jaynes não podem ser considerados tendo *zero* influência.

No entanto, essa pequena força da autoridade é apenas *ceteris paribus* e pode ser facilmente superada por argumentos mais fortes. Eu mesmo identifiquei uma pequena errata em um dos livros de Jaynes, pois a álgebra é mais poderosa que a autoridade.

## Referências

[1] Pearl, Probabilistic Reasoning in Intelligent Systems.

[2] Judea Pearl, Causality: Models, Reasoning, and Inference, 2nd ed. (New York: Cambridge University Press, 2009).

## 64 - Abraça a pergunta



Na arte da racionalidade, há uma disciplina de proximidade com o problema — trata-se de tentar examinar evidências que estejam o mais próximo possível da questão original, de modo a excluir o máximo de outros argumentos quanto for possível.

Os irmãos Wright dizem: “Meu avião vai voar”. Se você observar a autoridade deles (mecânicos de bicicletas que são excelentes físicos amadores) e, então, fizer uma comparação com a autoridade de, digamos, Lorde Kelvin, você descobrirá que Lorde Kelvin é a maior autoridade.

No entanto, se você exige ver os cálculos dos irmãos Wright e consegue acompanhá-los, e exige ver os cálculos de Lorde Kelvin (que provavelmente não possui nenhum além de sua própria incredulidade), então a autoridade perde relevância.

Ao observar o avião voar de verdade, os próprios cálculos se tornam questionáveis para muitos propósitos, e a autoridade de Kelvin nem vale a pena ser considerada.

Quanto mais diretamente seus argumentos abordarem uma questão, sem inferências intermediárias — quanto mais próximos os nós observados estiverem do nó consultado, na Grande Teia de Causalidade — mais poderosa será a evidência. É um teorema desses diagramas causais que você nunca pode obter mais informações de nós distantes do que de nós estritamente mais próximos que [bloqueiam os distantes](#).

Jerry Cleaver disse: “O que te deixa para baixo não é a falha em aplicar alguma técnica complexa de alto nível. É negligenciar os fundamentos. Não ficar atento<sup>18</sup>.” [\[1\]](#)

Assim como é melhor argumentar a física do que as credenciais, é também melhor argumentar a física do que a racionalidade. Quem era mais racional, os irmãos Wright ou Lorde Kelvin? Se pudermos verificar seus cálculos, não precisamos nos preocupar! A virtude de um racionalista não pode fazer um avião voar diretamente.

Se você esquecer esse princípio, [aprender sobre mais vieses pode te machucar](#), pois o distrairá de argumentos mais diretos. É muito fácil argumentar que alguém está exibindo o viés n.º 182 em seu repertório de acusações totalmente genéricas, mas você não pode resolver uma questão factual sem evidências mais precisas. [Se houver razões tendenciosas para dizer que o Sol está brilhando, isso não o torna escuro](#).

Assim como [você nem sempre pode vivenciar o hoje](#), nem sempre pode verificar os cálculos hoje.<sup>2</sup> Às vezes, você não conhece material de apoio suficiente, às vezes há informações privadas, às vezes simplesmente não há tempo. Há um número infelizmente grande de vezes em que vale a pena julgar a racionalidade do orador. Você deve sempre o fazer com um sentimento de vazio em seu coração, porém, uma sensação de que algo está faltando.

Sempre que possível, aproxime-se da pergunta original — pressione-se contra ela — aproxime-se o suficiente para *abraçá-la*!

---

18 NT: Texto original em inglês. *What does you in is not failure to apply some high-level, intricate, complicated technique. It's overlooking the basics. Not keeping your eye on the ball.*

## Referências

[1] Jerry Cleaver, *Immediate Fiction: A Complete Writing Course* (Macmillan, 2004).



## 65 - A Racionalidade e a língua inglesa



Em resposta à minha discussão sobre [Luzes de aplauso](#), alguém mencionou que minha escrita lembrava [Politics and the English Language](#) (Política e Língua Inglesa) [1], de George Orwell. Fiquei honrado com a comparação, especialmente porque já havia pensado nesse assunto para hoje.

Se você [realmente deseja ter a perspectiva de um artista](#) sobre a racionalidade, leia Orwell; ele é leitura essencial tanto para racionalistas quanto para escritores. Orwell não era um cientista, mas um escritor; suas ferramentas eram as palavras, não os números; seu adversário não era a natureza, mas o mal humano. Se você quer manter pessoas presas por anos sem julgamento, precisa encontrar uma maneira diferente de expressar isso, em vez de simplesmente dizer: “Vou prender o Sr. Jennings por anos sem julgamento”. Você precisa obscurecer o pensamento do ouvinte, evitando que imagens claras choquem a consciência. Você diria: “Elementos não confiáveis foram submetidos a um processo de justiça alternativo”.

Orwell era um oponente indignado do totalitarismo e do pensamento obscurantista, no qual o mal se esconde. Foi assim que os escritos de Orwell sobre a linguagem se tornaram documentos clássicos para os racionalistas, comparáveis a Feynman, Sagan ou Dawkins.

“Os escritores são instruídos a evitar o uso da voz passiva”. Um racionalista cuja formação é exclusivamente científica pode não perceber a falha na frase anterior; no entanto, qualquer pessoa que tenha escrito um pouco logo a notará.

Escrevi a frase na voz passiva, sem mencionar quem instrui os escritores a evitar o uso dela. A voz passiva remove o agente da ação, deixando apenas o sujeito da ação. “Elementos não confiáveis foram submetidos a um processo de justiça alternativo” — submetidos por quem? O que constitui um “processo de justiça alternativo”? Com frases substantivas estáticas suficientes, é possível evitar que qualquer coisa desagradável aconteça de fato.

É verdade que muitos artigos de periódicos são escritos em voz passiva. (Perdoe-me, alguns cientistas escrevem seus artigos de periódicos em voz passiva. Não é como se os artigos estivessem sendo escritos por ninguém, sem ninguém para culpar.) Parece mais autoritário dizer “Os indivíduos receberam *Progenitorivox*” do que “Eu dei a cada estudante universitário um frasco de 20 *Progenitorivox* e disse-lhes para tomar um todas as noites até que fossem embora.” Ao remover o cientista da descrição, apenas os dados mais importantes permanecem. No entanto, na realidade, o cientista está presente, os sujeitos são estudantes universitários e o *Progenitorivox* não foi “administrado”, mas entregue com instruções. A voz passiva obscurece a realidade.

Com base nos comentários que recebo, alguém poderia protestar que usar a voz passiva em um artigo de periódico dificilmente é um pecado, afinal, se você pensar sobre isso, pode perceber que o cientista está presente. Parece não haver uma falha lógica. É por isso que os racionalistas precisam ler Orwell, não apenas Feynman ou mesmo Jaynes. A não-ficção transmite conhecimento, enquanto a ficção transmite experiência. A ciência médica pode extrapolar o que aconteceria a um ser humano desprotegido no vácuo.

Mas a ficção pode fazer você vivenciar isso.

Alguns racionalistas tentarão analisar uma frase enganosa, tentarão ver se pode haver algo significativo nela, tentarão construir uma interpretação lógica. Eles serão caridosos, darão ao autor o benefício da dúvida. No entanto, os autores são treinados para não se dar o benefício da dúvida. O que quer que o público pense que você disse é o que você disse, quer você realmente quisesse dizer isso ou não; você não pode discutir com o público, não importa o quão inteligentes sejam suas justificativas.

Um escritor sabe que os leitores não param um minuto para pensar. Uma experiência ficcional é um fluxo contínuo de primeiras impressões. Um escritor-racionalista presta atenção à experiência que as palavras criam. Se você está avaliando a racionalidade pública de uma declaração e analisando as palavras deliberadamente, reformulando proposições, experimentando diferentes significados, buscando pepitas de veracidade, então você está perdendo a noção da primeira impressão — o que o público vê, ou melhor, sente.

Entendo seu ponto de vista em relação à importância de criar experiências vívidas e impactantes por meio das palavras. Um romancista notaria certamente o erro na frase “Os sujeitos receberam *Progenitorivox*”. Essa frase não oferece vida ou experiência para o leitor vivenciar. Ela apenas cria uma sensação distante de autoridade, e isso é tudo — a única experiência é a sensação de ouvir algo confiável. Um romancista poderia mostrar, por meio de substantivos mais concretos, o que realmente aconteceu — o pós-doutorando com a garrafa na mão, tentando parecer severo, o aluno ouvindo com um sorriso nervoso.

Meu ponto não é dizer que os artigos de periódicos devem ser escritos como romances, mas sim que um racionalista deve estar ciente das experiências que as palavras criam. Um racionalista deve compreender a mente e como operá-la. Isso inclui o fluxo de consciência, a parte de si mesmo que se desenvolve na linguagem. Um racionalista deve estar consciente do impacto real e experimental das frases, além de sua mera semântica proposicional.

Em resumo, o significado não desculpa o impacto!

Não me importo com a interpretação racional que você pode construir para uma luz de aplausos como “A IA deve ser desenvolvida por meio de processos democráticos”. Isso não pode desculpar seu impacto irracional de sinalizar ao público para aplaudir, sem mencionar sua imprecisão nebulosa que levanta questionamentos.

Aqui está Orwell, protestando contra o impacto dos clichês, e seu efeito na experiência do pensamento:

“Quando alguém assiste a algum discurso cansativo na plataforma, repetindo mecanicamente frases familiares como “Bestial”, “Atrocidades”, “Salto de Ferro”, “Tiranía Manchada de Sangue”, “Povos Livres do Mundo” e “Ficar Ombro a Ombro”, muitas vezes temos a curiosa sensação de que não estamos presenciando um ser humano vivo, mas apenas uma espécie de manequim. Um orador que utiliza essa forma de fraseologia percorreu um certo caminho rumo à transformação em uma máquina. Os ruídos apropriados saem de sua laringe, mas seu cérebro não está envolvido, como estaria se ele estivesse escolhendo suas palavras por si mesmo.

Acima de tudo, é fundamental permitir que o significado escolha a palavra, e não o contrário. Na prosa, render-se às palavras é a pior coisa que se pode fazer com elas. Quando pensamos em um objeto concreto, o fazemos sem palavras e, caso queiramos descrever aquilo que visualizamos, buscamos as palavras exatas que parecem se adequar à imagem. Ao pensar em algo abstrato, tendemos a recorrer às palavras desde o início, a menos que façamos um esforço consciente para evitá-lo. Nesse caso, o dialeto comum assume o controle, resultando em confusão ou até mesmo alteração de seu significado. É melhor adiar o uso de palavras tanto quanto possível e buscar obter o significado mais claro por meio de imagens e sensações”<sup>19</sup>.

Charles Sanders Peirce poderia ter escrito esse último parágrafo. Existem diversas rotas que podem levar ao *Caminho*.

---

19 NT: Texto original em inglês. *When one watches some tired hack on the platform mechanically repeating the familiar phrases—bestial, atrocities, iron heel, bloodstained tyranny, free peoples of the world, stand shoulder to shoulder—one often has a curious feeling that one is not watching a live human being but some kind of dummy. . . A speaker who uses that kind of phraseology has gone some distance toward turning himself into a machine. The appropriate noises are coming out of his larynx, but his brain is not involved, as it would be if he were choosing his words for himself. . . What is above all needed is to let the meaning choose the word, and not the other way around. In prose, the worst thing one can do with words is surrender to them. When you think of a concrete object, you think wordlessly, and then, if you want to describe the thing you have been visualising you probably hunt about until you find the exact words that seem to fit it. When you think of something abstract you are more inclined to use words from the start, and unless you make a conscious effort to prevent it, the existing dialect will come rushing in and do the job for you, at the expense of blurring or even changing your meaning. Probably it is better to put off using words as long as possible and get one’s meaning as clear as one can through pictures and sensations.*

## Referências

[1] George Orwell, "Politics and the English Language," Horizon (April 1946).



**G - Contra a racionalização**



## 66 - Maldade humana e pensamento confuso



[George Orwell](#) testemunhou a queda do mundo civilizado para o totalitarismo, a conversão e corrupção de país após país, a bota pisando eternamente no rosto humano. Lembre-se, é eterno. Você nasceu tarde demais para recordar uma época em que o avanço do totalitarismo parecia imparável, quando um país após o outro sucumbia à polícia secreta e à batida da meia-noite, enquanto os acadêmicos de universidades livres aplaudiam os expurgos da União Soviética como progresso. Parece tão distante para você quanto a ficção, difícil de levar a sério. Pois em sua época, o Muro de Berlim caiu. E se o nome de Orwell não está gravado em uma dessas pedras, deveria estar.

Orwell vislumbrou o destino da humanidade e lutou vigorosamente para desviá-la de seu curso. Sua arma era a escrita clara. Orwell compreendia que uma linguagem confusa resulta em pensamentos confusos; sabia que a maldade humana e o pensamento confuso se entrelaçam como fios de DNA conjugados.<sup>[1]</sup>

Em nosso tempo, o discurso e a escrita política são, na maioria, uma defesa do indefensável. Questões como a perpetuação do domínio britânico na Índia, os expurgos e deportações russas, o lançamento de bombas atômicas no Japão podem, de fato, ser justificadas, mas somente por argumentos brutais demais para a maioria enfrentar e que não se alinham com os objetivos declarados dos partidos políticos. Portanto, a linguagem política consiste, na maioria, de eufemismos, petições de princípio e ambiguidades puras e nebulosas. Aldeias indefesas são bombardeadas pelo ar, os habitantes são expulsos para o campo, o gado é metralhado e as cabanas são incendiadas com balas incendiárias: isso é chamado de PACIFICAÇÃO<sup>20</sup>...

Orwell deixou claro o propósito de sua clareza:

Se você simplificar seu inglês, estará livre das piores insanidades da ortodoxia. Você não conseguirá falar nenhum dos dialetos necessários, e quando fizer um comentário estúpido, sua estupidez será óbvia, mesmo para você<sup>21</sup>.

Tornar nossa estupidez óbvia, até mesmo para nós mesmos — esse é o cerne do *Overcoming Bias*.

O mal se esgueira nas sombras escuras da mente. Olhamos para trás com a clareza da história e choramos ao lembrar das fomes planejadas por Stalin e Mao, que tiraram a vida de dezenas de milhões. Chamamos isso de mal porque foi feito pela intenção humana deliberada de infligir dor e morte a seres humanos inocentes. Chamamos isso de mal devido à repulsa que sentimos contra ele, olhando para trás com a clareza da história. Para que os perpetradores do mal evitem a oposição natural, a repulsa deve permanecer adormecida. A clareza deve ser evitada a todo custo. Assim como os humanos de visão clara tendem a se opor ao mal que veem, o mal humano, onde quer que exista, começa a turvar o pensamento.

---

20 NT: Texto original em inglês. *In our time, political speech and writing are largely the defence of the indefensible. Things like the continuance of British rule in India, the Russian purges and deportations, the dropping of the atom bombs on Japan, can indeed be defended, but only by arguments which are too brutal for most people to face, and which do not square with the professed aims of the political parties. Thus political language has to consist largely of euphemism, question-begging and sheer cloudy vagueness. Defenceless villages are bombarded from the air, the inhabitants driven out into the countryside, the cattle machine-gunned, the huts set on fire with incendiary bullets: this is called PACIFICATION...*

21 NT: Texto original em inglês. *If you simplify your English, you are freed from the worst follies of orthodoxy. You cannot speak any of the necessary dialects, and when you make a stupid remark its stupidity will be obvious, even to yourself.*

A obra *1984* expõe isso claramente: os maiores vilões de Orwell são os manipuladores e retocadores de fotografias (baseados nas manipulações e retoques históricos na União Soviética). No auge de toda a escuridão no Ministério do Amor, O'Brien tortura Winston para fazer com que ele admita que dois mais dois são cinco [2]:

“Você se lembra”, continuou ele, “de ter escrito em seu diário: ‘Liberdade é a liberdade de dizer que dois mais dois são quatro?’”

“Sim”, disse Winston.

O'Brien levantou a mão esquerda, de costas para Winston, com o polegar escondido e os quatro dedos estendidos.

“Quantos dedos estou segurando, Winston?”

“Quatro.”

“E se o partido disser que não são quatro, mas cinco, então quantos são?”

“Quatro.”

A palavra escapou de seus lábios em um suspiro doloroso. O ponteiro do mostrador saltou para cinquenta e cinco. O suor brotou por todo o corpo de Winston. O ar encheu seus pulmões e saiu em gemidos profundos que ele não conseguia conter, mesmo cerrando os dentes. O'Brien observou-o, com os quatro dedos ainda estendidos. Ele puxou a alavanca para trás. Desta vez, a dor foi apenas ligeiramente amenizada.<sup>22</sup>

Fico constantemente horrorizado com pessoas aparentemente inteligentes, como o colega de Robin Hanson, [Tyler Cowen](#), que não consideram importante superar o viés. Estamos falando da sua mente. Sua inteligência humana. Isso o diferencia de um macaco. Você não acha que o funcionamento da mente é importante? Você não acha que as disfunções sistemáticas da mente são relevantes? Você acredita que a Inquisição teria torturado bruxas se todas fossem idealmente racionais?

Tyler Cowen aparentemente acredita que superar o viés é tão tendencioso quanto ter viés: “Eu vejo o blog de Robin como um exemplo de viés e, de fato, mostra que o viés pode ser muito útil”. Espero sinceramente que isso seja apenas o resultado de um pensamento excessivamente abstrato na tentativa de parecer inteligente. Tyler realmente acredita que a indiferença em relação ao valor da vida humana é equivalente a tentar criar planos que realmente salvem o maior número possível de vidas?

Orwell foi obrigado a lutar contra uma atitude semelhante — admitir qualquer distinção era considerado ingenuidade juvenil:

Stuart Chase e outros chegaram perto de afirmar que todas as palavras abstratas são desprovidas de sentido, usando isso como pretexto para defender uma espécie de quietismo político. Se você não sabe o que é o fascismo, como pode lutar contra o fascismo?<sup>23</sup>

Talvez superar o viés não pareça suficientemente empolgante quando enquadrado como uma luta contra simples erros acidentais. Pode ser que seja mais difícil se entusiasmar quando não há um mal claro

---

22 NT: Texto original em inglês. “Do you remember,” he went on, “writing in your diary, ‘Freedom is the freedom to say that two plus two make four?’” “Yes,” said Winston. O'Brien held up his left hand, its back towards Winston, with the thumb hidden and the four fingers extended. “How many fingers am I holding up, Winston?” “Four.” “And if the party says that it is not four but five—then how many?” “Four.” The word ended in a gasp of pain. The needle of the dial had shot up to fifty-five. The sweat had sprung out all over Winston's body. The air tore into his lungs and issued again in deep groans which even by clenching his teeth he could not stop. O'Brien watched him, the four fingers still extended. He drew back the lever. This time the pain was only slightly eased.

23 NT: Texto original em inglês. *Stuart Chase and others have come near to claiming that all abstract words are meaningless, and have used this as a pretext for advocating a kind of political quietism. Since you don't know what Fascism is, how can you struggle against Fascism?*

para combater. Portanto, deixemos absolutamente claro que onde houver maldade humana no mundo, onde houver crueldade, tortura e assassinato deliberado, há preconceitos envolvidos. Onde pessoas de visão clara se opõem a esses preconceitos, o mal oculto contra-ataca. A verdade tem inimigos. Se *Overcoming Bias* fosse um boletim informativo na antiga União Soviética, cada postagem e comentário do *Overcoming Bias* teria sido enviado para campos de trabalho.

Ao longo da história humana, todo avanço foi impulsionado por uma nova clareza de pensamento. Com exceção de algumas catástrofes naturais, todo grande infortúnio foi motivado por estupidez. Nosso último inimigo somos nós mesmos; esta é uma guerra e nós somos soldados.

## Referências

[1] Orwell, "Politics and the English Language."

[2] George Orwell, 1984 (Signet Classic, 1950).

## 67 - Saber sobre vieses pode machucar as pessoas



Uma vez, tentei explicar o problema da calibração especializada para a minha mãe, dizendo: “Então, quando um especialista diz que está 99% confiante, isso só acontece cerca de 70% das vezes”. Houve uma pausa, e de repente percebi estar falando com a minha mãe, então apressei-me em acrescentar: “É claro que você precisa aplicar esse ceticismo de maneira imparcial, inclusive a si mesma, em vez de usá-lo apenas para argumentar contra qualquer coisa com a qual você discorde...”

E minha mãe disse: “Você está brincando? Isso é ótimo! Vou usar sempre!”

O *Motivated skepticism in the evaluation of political Beliefs* (Ceticismo Motivado na Avaliação de Crenças Políticas) de Taber e Lodge descreve a confirmação de seis previsões [\[1\]](#):

1. Efeito de atitude prévia. Indivíduos que possuem um forte sentimento sobre um determinado assunto — mesmo quando incentivados a serem objetivos — tendem a avaliar argumentos que apoiam sua opinião de forma mais favorável do que argumentos contrários.
2. Viés de desconirmação. Os sujeitos gastarão mais tempo e recursos cognitivos subestimando argumentos contrários do que argumentos a favor.
3. Viés de confirmação. Indivíduos que têm a liberdade de escolher suas fontes de informação tendem a buscar fontes que apoiam suas crenças, em vez de fontes que as contradizem.
4. **Polarização de atitudes. Apresentar um conjunto aparentemente equilibrado de argumentos a favor e contra um assunto pode acentuar a polarização inicial dos indivíduos.**
5. Efeito de força de atitude. Sujeitos que expressam atitudes mais fortes serão mais propensos aos vieses acima.
6. **Efeito de sofisticação. Sujeitos com conhecimento político, por possuírem maior munção para contra-argumentar fatos e argumentos incongruentes, estarão mais propensos aos vieses acima.**

Se você já é irracional desde o início, ter *mais* conhecimento pode prejudicá-lo. Embora para um verdadeiro bayesiano, a informação nunca teria utilidade esperada negativa. No entanto, os humanos não são perfeitos em lidar com a Teoria de Bayes. Se não tivermos cuidado, podemos nos prejudicar.

Já vi pessoas profundamente confusas devido ao próprio conhecimento de seus vieses. Elas têm mais recursos para argumentar contra qualquer coisa que não gostem. E esse problema — ter muitos recursos prontos — é uma das principais formas em que pessoas com alta capacidade mental acabam sendo burras, no sentido de “disracionalidade” de Stanovich.

Você conhece pessoas que se encaixam nessa descrição? Pessoas com um alto fator  $g$  que acabam sendo menos eficazes por serem muito sofisticadas como argumentadoras? Você acredita que poderia ajudá-las — tornando-as racionalistas mais eficazes — se simplesmente lhes apresentasse uma lista de vieses clássicos?

Lembro-me de alguém que aprendeu sobre o problema de calibração/excesso de confiança. Logo depois ele disse: “Bem, você não pode confiar em especialistas; eles estão errados com tanta frequência — como os experimentos mostraram. Então, quando prevejo o futuro, prefiro presumir que as coisas continuarão historicamente como sempre foram...” e partiu para toda essa extrapolação complexa, propensa a erros e altamente questionável. Por algum motivo, quando se tratava de confiar em suas próprias conclusões preferidas, todos esses vieses e falácias pareciam muito menos *importantes* — saltavam muito menos à mente — do que quando ele precisava contra-argumentar com alguém.



Contei a ele sobre o problema do viés de confirmação e de argumento sofisticado e, então, na próxima vez que eu disse algo de que ele não gostou, ele me acusou de ser um argumentador sofisticado. Ele não tentou apontar nenhum argumento sofisticado em particular, nenhuma falha em particular — apenas balançou a cabeça e suspirou tristemente sobre como eu aparentemente estava usando minha própria inteligência para me derrotar. Ele havia adquirido ainda outro contra-argumento completamente geral.

Até mesmo a ideia de um “argumentador sofisticado” pode ser perigosa se for facilmente associada a alguém aparentemente inteligente que diz algo que você não gosta.

Como alguém que visa aprender com seus erros, na minha última apresentação sobre heurísticas e vieses, abordei a falácia da conjunção e a heurística da representatividade, antes de abordar viés de confirmação, viés de desconfirmação, argumento sofisticado, ceticismo motivado e outros efeitos de atitudes de diferentes perspectivas. Passei os próximos trinta minutos *insistindo* nesse tema, reintroduzindo-o a partir de quantas perspectivas diferentes eu pudesse.

Eu queria despertar o interesse do meu público pelo assunto. Bem, uma simples descrição da falácia da conjunção e da representatividade seria suficiente para isso. Mas e se eles realmente se interessassem? O material sobre viés é principalmente da psicologia cognitiva devido à própria psicologia cognitiva. Durante aquela palestra, eu tinha que alertar meu público sobre as consequências do viés, caso contrário, provavelmente não ouviriam nada sobre isso.

Quer eu faça isso no papel ou em discurso, agora tento nunca mencionar calibração e excesso de confiança, a menos que tenha falado primeiro sobre viés de desconfirmação, ceticismo motivado, argumentadores sofisticados e disracionalidade no mentalmente ágil.

Primeiro, não faça nenhum mal!

## Referências

[1] Charles S. Taber and Milton Lodge, “Motivated Skepticism in the Evaluation of Political Beliefs,” *American Journal of Political Science* 50, no. 3 (2006): 755–769, doi:10.1111/j.1540-5907.2006.00214.x.

## 68 - Atualize-se progressivamente



[A política é o assassino da mente](#). Debate é guerra, [argumentos são soldados](#). Existe a tentação de buscar maneiras de interpretar todos os resultados experimentais possíveis para confirmar sua teoria, como proteger uma cidadela contra todos os possíveis ataques. Mas isso é algo que você não pode fazer. É matematicamente impossível. Para cada expectativa de evidência, há uma expectativa igual e oposta de contra evidência.

Contudo, não há problema se sua crença estimada não for perfeitamente defendida. Se a hipótese é que uma moeda cai cara 95% das vezes, então é esperado que, uma vez em vinte, você veja o que parece ser uma evidência contrária. Isso está certo. É normal. É até esperado, contanto que você tenha dezenove observações de apoio para cada uma contrária. Um modelo probabilístico pode [receber um ou dois golpes](#) e ainda sobreviver, desde que esses golpes não continuem chegando.

No entanto, acredita-se amplamente, especialmente no tribunal da opinião pública, que uma teoria verdadeira não pode ter falhas e uma teoria falsa *não* pode ter sucessos.

Você encontra pessoas que apresentam apenas uma peça do que consideram como evidência e alegam que sua teoria pode “explicá-la”, como se isso fosse todo o apoio necessário para qualquer teoria. Aparentemente, uma teoria falsa não pode ter nenhuma evidência de apoio; é impossível que uma teoria falsa se encaixe em um único evento. Portanto, uma única evidência de confirmação é tudo que qualquer teoria precisa.

É apenas um pouco menos tolo apresentar uma única peça de contra-evidência probabilística como refutação, como se fosse impossível para uma teoria correta ter qualquer argumento contrário. No entanto, é assim que os seres humanos têm argumentado por eras e eras, tentando refutar todos os argumentos do inimigo, negando ao inimigo até mesmo um único pinga de apoio. As pessoas desejam que seus debates sejam unilaterais; estão acostumadas a um mundo onde suas teorias preferidas não encontram nenhuma resistência. Portanto, permitir um único elemento de contra-evidência probabilística seria o fim do mundo.

Simplesmente sei que alguém na plateia vai dizer: “Mas você não pode conceder nem um único ponto se quiser vencer debates no mundo real! Se você admitir que existem quaisquer contra-argumentos, o Inimigo vai insistir neles repetidamente — você não pode deixar o Inimigo fazer isso! Você vai perder! O que poderia ser mais visceralmente aterrorizante do que isso?”

Tanto faz. A racionalidade não é para vencer debates; é para decidir de que lado se posicionar. Se você já decidiu qual lado defender, o trabalho da racionalidade está concluído dentro de você, para o bem ou para o mal. Mas como você pode decidir de que lado argumentar? Se *escolher o lado errado* é visceralmente aterrorizante, mesmo que seja apenas um pouco, é melhor você integrar todas as evidências.

A racionalidade não é um passeio, mas uma dança. Em cada passo dessa dança, seu pé deve pousar exatamente no ponto correto, nem para a esquerda, nem para a direita. Aumentando a crença com cada pinga de evidência confirmatória. Diminuindo a crença com cada pinga de evidência contrária. Sim, *diminuindo*. Mesmo com um modelo correto, se não for um modelo preciso, às vezes você precisará revisar sua crença.

Se um ou dois pingos de evidência forem contrários à sua crença, tudo bem. Isso acontece às vezes com evidências probabilísticas para teorias não precisas. (Se uma teoria precisa falhar, você está em apuros!) Apenas reduza um pouco sua crença — a probabilidade, a razão de chances ou até mesmo um peso subjetivo

de credibilidade em sua mente. Apenas diminua um pouco e aguarde por mais evidências. Se a teoria for verdadeira, evidências de apoio surgirão em breve e a probabilidade aumentará novamente. Se a teoria for falsa, você realmente não deseja mantê-la de qualquer maneira.

O problema de usar um raciocínio qualitativo, preto e branco, binário, é que qualquer observação única pode destruir a teoria ou não. Quando nem mesmo uma única observação contrária é permitida, isso cria dissonância cognitiva e deve ser descartada com argumentos. Isso exclui o progresso incremental; exclui a correta integração de todas as evidências. Ao raciocinarmos de forma probabilística, percebemos que, em média, uma teoria correta gerará um peso maior de apoio do que de contraponto. Portanto, você pode dizer a si mesmo, sem medo: “Esta é uma evidência levemente contrária, vou diminuir a minha crença”. Sim, diminuir. Isso não destrói sua teoria estimada. Isso é raciocínio qualitativo; pense quantitativamente.

Para cada expectativa de evidência, há uma expectativa igual e oposta de contra-evidência. Em todas as ocasiões, você deve, em média, antecipar a revisão de suas crenças para baixo, assim como antecipa revisá-las para cima. Se você acha que já sabe quais evidências virão, então já deve estar bastante certo de sua teoria — probabilidade próxima de 1 - o que não deixa muito espaço para a probabilidade aumentar ainda mais. E, por mais improvável que pareça encontrar evidências que não confirmem, o deslocamento para baixo resultante deve ser grande o suficiente para equilibrar com precisão o ganho antecipado do outro lado. A média ponderada de sua probabilidade posterior esperada deve ser igual à probabilidade anterior.

Quão tolo é ter medo de revisar sua probabilidade, se você está realmente investigando um assunto? Em média, você deve antecipar tanto o deslocamento para baixo quanto o deslocamento para cima de cada observação individual.

Pode acontecer que um pequeno sinal de contraponto apareça repetidamente, enquanto o novo apoio demora a aparecer. Você pode descobrir que sua crença está diminuindo cada vez mais. Até que, finalmente, você perceba de que lado as evidências estão se acumulando contra você. Nesse momento de realização, não adianta inventar desculpas. Nesse momento de realização, você já abandonou sua crença estimada. Ótimo! É hora de comemorar! Abra uma garrafa de champanhe ou peça uma pizza! Afinal, você não se fortalece ao manter as crenças com as quais começou.

## 69 - Um argumento contra um exército



Mencionei um estilo de raciocínio no qual [nenhum argumento contrário é permitido, resultando na exclusão de qualquer observação não sustentada](#). Aqui, sugiro que, quando as pessoas se deparam com um argumento contrário, elas evitam diminuir sua confiança ao revisar o suporte já conhecido.

Suponhamos que o país de *Freedonia*<sup>24</sup> esteja debatendo se seu vizinho, *Sylvania*, é responsável por uma recente onda de ataques de meteoros em suas cidades. Existem várias evidências que sugerem isso: os meteoros atingiram cidades próximas à fronteira com *Sylvania*; houve atividade suspeita nos mercados de ações de *Sylvania* antes dos ataques; e o embaixador *Sylvaniano Trentino* foi ouvido mencionando “vingança celestial”.

Alguém se aproxima de você e diz: “Não acredito que *Sylvania* seja responsável pelos ataques de meteoros. Eles negociam bilhões de dinares conosco anualmente.” E você responde: “Bem, os meteoros atingiram cidades próximas a *Sylvania*, houve atividade suspeita em seu mercado de ações e o embaixador deles mencionou vingança celestial.” Ao enfatizar esses três argumentos em vez do primeiro, você mantém sua crença de que *Sylvania* é responsável — você acredita, em vez de descartar, qualitativamente. Claramente, o peso das evidências aponta contra *Sylvania*.

Então outra pessoa se aproxima de você e diz: “Não acredito que *Sylvania* seja responsável pelos ataques de meteoros. Realizar um ataque com asteroides é realmente difícil. *Sylvania* nem tem um programa espacial.” Você responde: “Mas os meteoros atingiram cidades próximas a *Sylvania*, e seus investidores sabiam disso, e o embaixador veio imediatamente e admitiu!” Novamente, esses três argumentos superam o primeiro (três argumentos contra um), então você mantém sua crença de que *Sylvania* é responsável.

De fato, suas convicções se fortalecem. Agora, em duas ocasiões diferentes, você avaliou o peso das evidências e, em ambas as vezes, o peso indicava que *Sylvania* era a responsável, em uma proporção de 3 para 1.

Você encontra mais argumentos vindos dos defensores de *Sylvania* — repetidamente, centenas de vezes — mas a cada vez o novo argumento é facilmente derrotado por 3 a 1. E em todas essas ocasiões, você se sente cada vez mais confiante de que *Sylvania* é de fato a responsável, alterando sua crença inicial conforme o peso das evidências.

O problema, é claro, é que ao testar argumentos que você já conhecia, você está contando as evidências duas vezes. Isso seria um erro grave, mesmo se todas as evidências fossem contadas duas vezes. (Imagine um cientista que realiza um experimento com 50 sujeitos e não obtém resultados estatisticamente não significativos, então o cientista conta todos os dados duas vezes.)

No entanto, contar seletivamente apenas algumas evidências duas vezes é uma farsa completa. Lembro-me de ter visto um desenho animado quando era criança, onde um vilão dividia o saque usando o seguinte algoritmo: “Um para você, um para mim. Um para você, um-para mim. Um para você, um-para mim.”

Conforme enfatizei no último ensaio, mesmo que uma crença estimada seja *verdadeira*, um racionalista às vezes precisa reduzir a probabilidade dessa crença ao integrar *todas* as evidências. Sim, o equilíbrio do

---

24 NT: Nome criado pelo autor que alude à palavra *free*, livre, em inglês.

suporte ainda pode favorecer a crença estimada, mas é necessário ajustar a probabilidade para *baixo* — sim, para *baixo* — do que era antes de considerar a evidência contrária. Não faz sentido testar argumentos de apoio, pois eles já foram considerados.

Porém, parece que quando as pessoas são confrontadas com um novo contra-argumento, elas procuram justificativas para não diminuir sua confiança e, é claro, encontram argumentos de apoio que já conhecem. Eu mesmo tenho que estar constantemente vigilante para evitar fazer isso! Parece tão natural quanto usar um escudo para bloquear um golpe de espada. Com o tipo certo de raciocínio falho, um punhado de argumentos de apoio — ou mesmo um único argumento — pode resistir a um exército de contradições.

## 70 - O resultado final



Duas caixas seladas foram colocadas para leilão, a caixa A e a caixa B. Uma e somente uma dessas caixas contém um valioso diamante. Existem todo tipo de sinais e presságios indicando se uma caixa contém um diamante; mas eu não tenho nenhum sinal que eu saiba ser perfeitamente confiável. Há um carimbo azul em uma das caixas, por exemplo, e eu sei que as caixas que contém diamantes são mais propensas a mostrar um carimbo azul do que as caixas que estão vazias. Ou uma caixa tem uma superfície brilhante, e eu suspeito — mas não tenho certeza — que nenhuma caixa que contenha diamantes é brilhante.

Agora, suponha haver um arguidor astuto, segurando uma folha de papel, e ele diz aos donos da caixa A e da caixa B: “Dê um lance pelos meus serviços, e eu argumentarei que a caixa de quem ganhar contém o diamante, para que ela receba um preço mais alto.” Assim, os proprietários das caixas oferecem lances, e o proprietário da caixa B faz uma oferta mais alta, ganhando os serviços do argumentador astuto.

O arguidor astuto começa a organizar seus pensamentos. Primeiro, ele escreve “E, portanto, a caixa B contém o diamante!” no final de sua folha de papel. Então, no topo do papel, o argumentador astuto escreve: “A caixa B mostra um carimbo azul”, e abaixo dele, “A caixa A é brilhante”, e então, “A caixa B é mais leve que a caixa A”, e assim por diante através de muitos sinais e presságios; ainda assim, o argumentador astuto ignora todos os sinais que poderiam argumentar a favor da caixa A. E então o arguidor astuto vem até mim e lê o escrito na folha de papel: “A caixa B mostra um carimbo azul, e a caixa A é brilhante”, e assim por diante, até chegar a “e, portanto, a caixa B contém o diamante.”

Mas considere: quando o arguidor astuto registrou sua conclusão por escrito e usou tinta em seu papel, o vínculo probatório entre a tinta física e as caixas físicas foi estabelecido de maneira definitiva.

Pode ser útil visualizar uma coleção de mundos — as ramificações de Everett ou os [duplos de Tegmark](#) — nos quais existe uma frequência objetiva na qual a caixa A ou a caixa B contém um diamante. Também há alguma frequência objetiva no subconjunto “mundos com uma caixa brilhante A”, onde a caixa B contém o diamante; e alguma frequência objetiva em “mundos com a caixa A brilhante e a caixa B com o carimbo azul”, onde a caixa B contém o diamante.

A tinta no papel é moldada em formas e curvas estranhas, que parecem com este texto: “Portanto, a caixa B contém o diamante.” Se você fosse um falante alfabetizado em inglês, poderia se confundir e pensar que a forma da tinta significava de alguma maneira que a caixa B continha o diamante. Pessoas instruídas a dizer a cor das imagens impressas e mostradas a imagem **Verde** geralmente dizem “verde” em vez de “vermelho”. Ser analfabeto ajuda, pois você não é confundido pela forma da tinta.

Para nós, o verdadeiro significado de algo está na sua conexão com outras coisas. Considere novamente a coleção de mundos, ramificações de Everett ou os duplos de Tegmark. No momento em que todos os argumentadores astutos em todos os mundos escrevem no papel, vamos supor que isso aconteça em um único momento — a correlação da tinta com as caixas fica estabelecida. O arguidor astuto escreve com caneta de tinta permanente; a tinta não mudará. As caixas também não mudarão. No subconjunto de mundos no qual a tinta diz: “Portanto, a caixa B contém o diamante”, já existe uma porcentagem fixa de mundos nos quais a caixa A contém o diamante. Isso não mudará, não importa o que seja escrito nas linhas em branco acima.

Dessa forma, o envolvimento probatório da tinta fica fixado e cabe a você decidir o que isso pode significar. Talvez os proprietários das caixas que acreditam ter um argumento mais forte sejam mais propensos a contratar publicitários, ou talvez os proprietários que temem suas próprias deficiências ofereçam lances mais

altos. Se os proprietários das caixas não entenderem os sinais e presságios, a tinta estará completamente desvinculada do conteúdo das caixas, embora possa fornecer informações sobre as finanças e os hábitos de licitação dos proprietários.

Agora, suponha que outra pessoa esteja verdadeiramente curiosa. Primeiramente, ela anota todos os sinais distintivos de ambas as caixas em um pedaço de papel e, em seguida, utiliza seu conhecimento e as leis da probabilidade para escrever na parte inferior: “Portanto, estimo uma probabilidade de 85% de que a caixa B contenha o diamante.” Ao examinar a cadeia de causa e efeito que levou a esta tinta física no papel físico, percebe-se que a cadeia de causalidade é influenciada por todos os sinais e presságios das caixas e depende deles. Isso significa que, em mundos com sinais e presságios diferentes, uma probabilidade diferente seria escrita na parte inferior da folha.

Assim sendo, a caligrafia do investigador curioso está entrelaçada com os sinais, presságios e conteúdo das caixas, enquanto a caligrafia do arguidor astuto é evidência somente de qual proprietário fez o lance mais alto. Há uma grande discrepância entre as indicações fornecidas pela tinta, apesar de alguém que lesse com desatenção as formas da escrita poder pensar que as palavras em inglês soam semelhantes.

Sua eficácia como racionalista é determinada por qualquer algoritmo que realmente escreva o resultado final de seus pensamentos. Por exemplo, se o seu carro faz um som metálico quando você freia, e você não quer arcar com os custos de substituir os freios, você pode procurar justificativas para evitar o reparo. Entretanto, a verdadeira medida de quão racional você é, em termos de ramificações de Everett ou mundos de Tegmark — que podemos considerar como uma descrição da sua efetividade como racionalista — é determinada pelo algoritmo *que* determina qual conclusão você buscará argumentos. Neste caso, o verdadeiro algoritmo é “Nunca conserte nada caro.” Se esse for um bom algoritmo, ótimo; se não for, paciência. De qualquer forma, os argumentos que você elabora acima da linha final não terão impacto algum.

Essa é uma advertência para que você reflita sobre o seu próprio pensamento, não uma refutação geral contra conclusões que você não gosta. De fato, é uma estratégia inteligente dizer “Meu oponente é um arguidor astuto” se você estiver se esforçando para manter as crenças que tinha originalmente. Mesmo o arguidor mais astuto do mundo pode apontar que o sol está brilhando e isso não mudará que provavelmente ainda é dia.

## 71 - Que evidências filtraram evidências?



Discuti o dilema do arguidor astuto, contratado para vender uma caixa que pode ou não conter um diamante. O arguidor astuto aponta que a caixa tem um selo azul, e é um fato válido e conhecido que caixas contendo diamantes são mais propensas do que caixas vazias a ostentar um selo azul. O que acontece neste ponto, de uma perspectiva bayesiana? Você deve, impotente, atualizar suas probabilidades, como o arguidor astuto assim deseja que você o faça?

Se você puder olhar para a caixa você mesmo, pode somar todas as evidências. Mas e se você não puder olhar? E se a única evidência que você tiver for a palavra do arguidor astuto, o qual é legalmente obrigado a fazer apenas declarações verdadeiras, mas não lhe diz tudo o que sabe? Cada declaração que o arguidor esperto faz é uma evidência válida — como você poderia não atualizar suas probabilidades? Deixa de ser verdade que, em uma determinada proporção de universos paralelos de Everett ou cópias do multiverso de Tegmark nos quais a caixa B tem um selo azul, a caixa B contém um diamante? De acordo com Jaynes, um bayesiano deve sempre considerar todas as evidências conhecidas, sob pena de paradoxo. Então, o arguidor astuto pode fazer você acreditar em qualquer coisa que ele escolher, se houver uma variedade suficiente de sinais para relatar seletivamente. Isso não parece correto.

Considere um caso mais simples, uma moeda viciada, que pode cair com cara em  $\frac{2}{3}$  [dois terços] das vezes e coroa em  $\frac{1}{3}$  [um terço] das vezes, ou cara em  $\frac{1}{3}$  [um terço] das vezes e coroa em  $\frac{2}{3}$  [dois terços] das vezes, sendo ambos os casos igualmente prováveis *a priori*, isto é, antes do lançamento. A cada resultado de Cara observado é 1 bit de evidência para uma moeda viciada para cara. Lanço a moeda dez vezes e digo a você: “A 4ª, 6ª e 9ª jogadas deram cara”. Qual é a sua probabilidade posterior de que a moeda esteja viciada para cara?

E a resposta é que pode ser quase qualquer coisa, dependendo de qual cadeia de causa e efeito está por trás da minha declaração — minha seleção de quais resultados relatar.

- Posso estar seguindo o algoritmo de relatar o resultado da 4ª, 6ª e 9ª jogadas, independentemente dos resultados dessas e de todas as outras jogadas. Se você sabe que estou usando esse algoritmo, as probabilidades posteriores são de 8:1 [oito em uma] a favor da moeda estar viciada para cara.
- Eu poderia estar relatando todos os lançamentos, e apenas os lançamentos, que deram cara. Nesse caso, você sabe que os outros 7 lançamentos deram coroa e as probabilidades posteriores são de 1:16 [uma em dezesseis] contra a moeda estar viciada para cara.
- Eu poderia ter decidido antecipadamente relatar o resultado da 4ª, 6ª e 9ª jogadas somente se a probabilidade da moeda estar viciada para cara excedesse 98%. E assim por diante.

Ou considere o problema de Monty Hall:

Em um programa de televisão, você tem a opção de escolher entre três portas que levam a três quartos. Você sabe que um dos quartos contém US\$100.000 e os outros dois estão vazios. O apresentador pede que você escolha uma porta e você escolhe a porta número 1. Em seguida, o apresentador abre a porta número 2, revelando um quarto vazio. Você gostaria de mudar para a porta número 3 ou ficar com a porta número 1?

A resposta depende do algoritmo do apresentador. Se o apresentador sempre abre uma porta que leva a um quarto vazio, você deve mudar para a porta número 3. Se o apresentador sempre abre a porta número 2, independentemente do que esteja por trás dela, as portas número 1 e número 3 têm uma proba-



bilidade de 50% de conter o dinheiro. Se o apresentador abrir apenas uma porta, caso você tenha escolhido inicialmente a porta com o dinheiro, então você deve ficar definitivamente com a porta número 1.

Você não deve apenas considerar que a porta número 2 está vazia, mas sim esse fato juntamente com o fato de o apresentador escolher abrir a porta número 2. Muitas pessoas ficam confusas com o problema clássico de Monty Hall porque elas atualizam apenas sua crença quando a porta número 2 está vazia, o que leva a pensar que as portas número 1 e número 3 têm probabilidades iguais de conter o dinheiro. É por isso que os bayesianos são instruídos a considerar todo o seu conhecimento, sob pena de paradoxo.

Quando alguém diz: “A 4ª jogada de moeda deu cara”, não estamos condicionando ao fato de a 4ª jogada de moeda ter dado cara—não estamos considerando o subconjunto de todos os mundos possíveis onde a 4ª jogada de moeda deu cara—em vez disso, estamos condicionando ao subconjunto de todos os mundos possíveis onde um falante seguindo um algoritmo particular disse “A 4ª jogada de moeda deu cara”. A frase falada não é o fato em si; não se deixe enganar pelos meros significados das palavras.

A maioria dos processos legais funciona com base na teoria de que [cada caso tem exatamente dois lados opostos](#) e que é mais fácil encontrar dois humanos tendenciosos do que um imparcial. Entre a acusação e a defesa, alguém tem motivação para apresentar qualquer evidência, então o tribunal verá todas as evidências; essa é a teoria. Se houver dois argumentadores inteligentes no dilema das caixas, não é tão bom quanto *um investigador curioso*, mas é quase tão bom. Mas isso é com duas caixas. A realidade frequentemente tem problemas com muitos lados, problemas profundos e respostas não óbvias, que não são facilmente encontradas por Azuis e Verdes gritando uns com os outros.

Cuidado para não abusar da noção de filtragem de evidências como um contra-argumento totalmente geral para excluir todas as evidências de que você não gosta: “Esse argumento foi filtrado, portanto, posso ignorá-lo”. Se você está irritado com um argumento contrário, então você está familiarizado com o caso e se importa o suficiente para tomar partido. Você provavelmente já conhece os argumentos mais fortes do seu próprio lado. Você não tem razão para inferir, de um argumento contrário, a existência de [novos](#) sinais e presságios favoráveis que você ainda não viu. Então você fica com os fatos desconfortáveis em si; um carimbo azul na caixa B ainda é evidência.

Mas se você está ouvindo um argumento pela primeira vez, e está ouvindo apenas um lado do argumento, então de fato você deve tomar cuidado! De certa forma, ninguém pode realmente confiar na teoria da seleção natural até após ter ouvido criacionistas por cinco minutos; e então eles sabem que é sólida.

## 72 - Racionalização



Em “[O resultado final](#)”, apresentei o dilema de duas caixas, sendo que apenas uma contém um diamante, com vários sinais e presságios como evidência. Contrastei o investigador curioso e o arguidor astuto. O investigador curioso anota todos os sinais e presságios, processa-os e finalmente escreve: “Assim, calculo uma probabilidade de 85% de que o diamante esteja na caixa B.” O arguidor astuto trabalha para o licitante mais alto e começa escrevendo: “Portanto, a caixa B contém o diamante” para depois selecionar sinais e presságios favoráveis e listá-los nas linhas acima.

O primeiro procedimento é a racionalidade. O segundo procedimento é geralmente conhecido como “racionalização”.

“Racionalização.” Que termo curioso... Eu diria que é uma palavra inadequada. Você não pode “racionalizar” o que ainda não é racional. É como se a “mentira” fosse chamada de “veracização.”

No nível puramente computacional, há uma diferença bastante grande entre:

1. Partir de evidências e, em seguida, processar os fluxos de probabilidade, a fim de produzir uma conclusão provável. (Isso equivale a escrever todos os sinais e presságios e, em seguida, avançar para uma probabilidade sobre o resultado final que depende desses sinais e presságios)
2. Começar com uma conclusão e, em seguida, analisar fluxos de probabilidade a fim de produzir evidências que aparentemente favoreçam essa conclusão. Esse método é questionável. Essa abordagem envolve anotar o resultado final primeiro e, em seguida, retroceder para selecionar [sinais](#) e presságios que serão apresentados nas linhas acima.

Quem foi o tolo que inventou palavras tão confusamente semelhantes, como “racionalidade” e “racionalização”, para descrever processos mentais tão extraordinariamente diferentes? Eu preferiria termos que tornassem a diferença algorítmica óbvia, como “racionalidade” contra “buraco negro cognitivo sugador gigante”.

Nem toda mudança é uma melhoria, mas toda melhoria é necessariamente uma mudança. Você não pode obter mais verdade para uma proposição fixa discutindo-a; você pode fazer mais pessoas acreditar em nela, mas não pode torná-la mais verdadeira. Para melhorar nossas crenças, devemos necessariamente mudá-las. Racionalidade é a operação que usamos para obter mais precisão em nossas crenças, alterando-as. A racionalização opera para fixar crenças no lugar; seria melhor chamá-la de “anti-racionalidade”, tanto por seus resultados pragmáticos quanto por seu algoritmo invertido.

“Racionalidade” é o fluxo *direto* que reúne evidências, pesa-as e produz uma conclusão. O inquiridor curioso usou um algoritmo de fluxo direto: primeiro reunindo as evidências, anotando uma lista de todos os sinais e presságios visíveis, que ele então processou para obter uma probabilidade anteriormente desconhecida da caixa contendo o diamante. Durante todo o tempo em que o processo de racionalidade decorreu, o inquiridor curioso ainda não sabia o seu destino, por isso estava curioso. No Caminho de Bayes, a probabilidade anterior é igual à probabilidade posterior esperada: se você conhece seu destino, já está lá.

A “racionalização” é um processo reverso, que parte da conclusão e vai em direção às evidências selecionadas. Primeiramente, é necessário anotar o resultado final, o qual é conhecido e fixo, e o objetivo do processo é descobrir quais argumentos devem ser anotados nas linhas anteriores. Isso, e não o resultado final, é a variável desconhecida durante a execução do processo.

Receio que a abordagem da Racionalidade Tradicional não conscientize adequadamente seus usuários sobre a diferença entre fluxo direto e fluxo reverso. Na Racionalidade Tradicional, não há nada de errado com o cientista que chega a uma hipótese preferida e depois busca um experimento que a comprove. Um racionalista tradicional aprovaria essa abordagem e diria: “Esse ímpeto é o motor que impulsiona o avanço da Ciência.” Bem, esse é o motor que impulsiona a Ciência. É mais fácil encontrar um acusador e um defensor, tendenciosos em direções opostas, do que encontrar um único ser humano imparcial.

Porém, apenas porque todos fazem uma determinada coisa, não significa que ela está certa. Seria ainda melhor se o cientista, ao chegar a uma hipótese preferida, decidisse testá-la por mera *curiosidade* — criando experimentos que levassem suas próprias crenças em uma direção desconhecida.

Se você realmente não sabe para onde está indo, provavelmente ficará bastante curioso sobre isso. A curiosidade é a primeira virtude, sem a qual seu questionamento será sem propósito e suas habilidades sem direção.

Sinta o fluxo da Força e certifique-se de que ele não está fluindo para trás.

## 73 - Um argumento racional



Você trabalha como gerente de campanha e acaba de ser contratado por *Mortimer Q. Snodgrass*, o candidato verde a prefeito de *Hadleyburg*. Enquanto lê um livro sobre racionalidade, uma pergunta surge em sua mente: “Como posso construir um argumento racional impecável de que *Mortimer Q. Snodgrass* é o melhor candidato a prefeito de *Hadleyburg*?”

Desculpe, mas isso não pode ser feito.

“O quê?”, você exclama. “E se eu usar apenas argumentos válidos para construir minha estrutura lógica? E se todos os fatos que eu mencionar forem verdadeiros, até onde sei, com evidências relevantes baseadas na Regra de Bayes?”

Desculpe, ainda assim não é possível. Você se derrota no momento em que especifica antecipadamente a conclusão do seu argumento.

Este ano, o jornal *Trombeta* de *Hadleyburg* enviou um questionário de 16 perguntas a todos os candidatos a prefeito, contendo questões como “Você consegue pintar com todas as cores do arco-íris?” e “Você tragou ou só fumou?”. Infelizmente, os escritórios do *Trombeta* foram destruídos por um meteorito antes da publicação. É lamentável, já que seu próprio candidato, *Mortimer Q. Snodgrass*, se destacou em 15 das 16 perguntas, em comparação com seus oponentes. O único ponto de discordância era a pergunta 11: “Você é, atualmente, ou já foi, um supervilão?”

Assim, você se sente tentado a publicar o questionário como parte de sua própria literatura de campanha... omitindo, é claro, a 11ª pergunta.

Isso cruza a linha entre racionalidade e racionalização. Os eleitores não podem mais se basear apenas nos fatos; eles devem considerar também como esses [fatos adicionais](#) são apresentados e inferir a existência de evidências ocultas.

Na verdade, você ultrapassou os limites ao considerar se o questionário era favorável ou desfavorável ao seu candidato antes de decidir se o publicaria ou não. “O quê!” você exclama. “Uma campanha deve divulgar fatos desfavoráveis ao seu candidato?” Mas coloque-se no lugar de um eleitor que continua tentando escolher um candidato — por que você censuraria informações úteis? Você não faria isso se estivesse genuinamente curioso. Se você estivesse fluindo da evidência para uma escolha desconhecida de candidato, em vez de fluir de um candidato fixo para determinar os argumentos.

Um argumento “lógico” é aquele que segue a partir de suas premissas. Portanto, o seguinte argumento é ilógico:

- Todos os retângulos são quadriláteros.
- Todos os quadrados são quadriláteros.
- Portanto, todos os quadrados são retângulos.

Esse silogismo não se torna lógico apenas pela verdade de suas premissas ou mesmo pela verdade de sua conclusão. Vale a pena distinguir deduções lógicas das ilógicas e recusar-se a desculpá-las, mesmo que suas conclusões sejam verdadeiras. Por um lado, essa distinção pode afetar como revisamos nossas crenças à luz de evidências futuras. Por outro lado, a negligência é um hábito.

Acima de tudo, o silogismo falha em apresentar a explicação real. Talvez todos os quadrados sejam retângulos, mas, se forem, isso não se deve ao fato de ambos serem quadriláteros. Podemos chamá-lo de um silogismo hipócrita — um que apresenta uma desconexão entre as razões declaradas e as razões reais.

Se você realmente deseja apresentar um argumento honesto e racional *para o seu candidato* durante uma campanha política, há apenas uma maneira de fazer isso:

- Antes de ser contratado por alguém, reúna todas as evidências possíveis sobre os diferentes candidatos.
- Crie uma lista de verificação que você mesmo utilizará para decidir qual candidato parece ser o melhor.
- Avalie cada ponto da lista de verificação.
- Opte pelo candidato que se destacar como vencedor.
- Se ofereça para ser o gerente de campanha desse candidato.
- Quando for solicitada a produção de materiais de campanha, imprima a sua lista de verificação.

Somente assim você poderá oferecer uma linha de raciocínio verdadeiramente racional, na qual [a conclusão](#) seja uma consequência direta das premissas anteriores. O que realmente determina sua conclusão deve ser o único conteúdo que você pode honestamente apresentar como justificativa.

## 74 - Evitando os verdadeiros pontos fracos de sua crença



Há alguns anos, minha bisavó faleceu, aos noventa anos, após uma longa, lenta e cruel desintegração. Nunca a conheci pessoalmente, mas na minha distante infância ela cozinhava para a família. Lembro-me de seu *gefilte fish*<sup>25</sup> e de seu rosto, e de que ela era gentil comigo. Em seu funeral, meu tio-avô, que cuidou dela por anos, falou. Ele disse, com a voz embargada, que Deus havia levado sua mãe lentamente: primeiro sua memória, depois sua capacidade de falar e, por último, seu sorriso. Quando Deus tirou finalmente o sorriso dela, ele soube que não demoraria muito para ela falecer, pois isso significava que ela estava quase completamente morta.

Ouvi isso e fiquei perplexo, porque era algo inconcebivelmente horrível de se acontecer a alguém e, portanto, não esperava que meu tio-avô atribuísse isso a Deus. Segundo a teologia judaica, Deus sustenta continuamente o universo e escolhe cada evento nele; mas normalmente, extrair implicações lógicas dessa crença é reservado para ocasiões mais felizes. Ao dizer “Deus fez isso!” apenas quando você foi abençoado com uma menina, e simplesmente não pensando “Deus fez isso!” para abortos, natimortos e mortes no berço, você pode criar uma imagem bastante distorcida da personalidade benevolente do seu Deus.

Por isso, fiquei surpreso ao ouvir meu tio-avô atribuir a lenta desintegração de sua mãe a um ato de Deus deliberado e estrategicamente planejado. Isso violou as regras da autoilusão religiosa como eu as entendia.

Se eu tivesse notado minha própria confusão, poderia ter feito uma previsão surpreendente e bem-sucedida. Não muito tempo depois, meu tio-avô deixou a religião judaica. (O único membro da minha família além de mim a fazê-lo, até onde eu sei.)

O [judaísmo ortodoxo moderno](#) é diferente de qualquer outra religião que já conheci e não sei como explicá-lo para alguém que não foi obrigado a estudar *Mishná* e *Gemara*<sup>26</sup>. Existe uma tradição de questionamento, mas é um tipo de questionamento diferente... Não seria nada surpreendente ouvir um rabino, durante o sermão semanal, apontar o conflito entre os sete dias da criação e os 13,7 bilhões de anos desde o Big Bang — porque ele achou que tinha uma explicação realmente inteligente para isso, envolvendo três outras referências bíblicas, um *Midrash*<sup>27</sup> e um artigo mal compreendido na *Scientific American*. No judaísmo ortodoxo, você pode perceber inconsistências e contradições, mas somente para explicá-las, e quem apresentar a explicação mais complicada ganha um prêmio.

Existe uma tradição de investigação. Mas você só ataca os alvos com o propósito de defendê-los. Você

---

25 NT: Original em iídiche, *gefilte fish*, ou peixe recheado com tradução livre para o português brasileiro, é um prato tradicional da culinária judaica ashkenazi (judeus da Europa Central e oriental) servido comumente nas festas judaicas, principalmente durante a festa de *Pessach*.

26 NT: Original em hebraico *Mishna* (משנה) em tradução livre para o português “repetição” ou “estudo”, e, do aramaico, *Gemara* (גמרא), em tradução livre para o português “conclusão” ou “complemento”. A *Mishná* e a *Gemara* são partes essenciais do Talmude judaico. A *Mishná*, compilada por volta de 200 EC, é o registro escrito da lei oral judaica. A *Gemara*, desenvolvida posteriormente, é um comentário detalhado sobre a *Mishná*. Juntas, formam o Talmude, um texto fundamental para a lei e filosofia judaica.

27 NT: Original em hebraico, *Midrash* (מדרש) em tradução livre para o português, significa “investigação” ou “interpretação”. *Midrash* é um método de interpretação e comentário bíblico na tradição judaica. Refere-se tanto ao processo de interpretação quanto à coleção de textos interpretativos que resultam desse processo. O *Midrash* busca explicar, expandir e aplicar as narrativas, leis e ensinamentos encontrados na Torá e em outros textos bíblicos.

só ataca alvos que sabe que pode defender.

No judaísmo ortodoxo moderno, não se enfatiza muito as virtudes da fé cega. É permitido duvidar. Mas não é permitido [ter sucesso na dúvida](#).

Espero que a grande maioria dos judeus ortodoxos educados tenha questionado sua fé em algum momento de suas vidas. Mas o questionamento provavelmente foi mais ou menos assim: “Segundo os céti- cos, a Torá diz que o universo foi criado em sete dias, o que não é cientificamente preciso. Mas será que as tribos originais de Israel, reunidas no Monte Sinai, conseguiriam entender a verdade científica, mesmo que tivesse sido apresentada a elas? Elas ao menos tinham uma palavra para ‘bilhões’? É mais fácil ver a história dos sete dias como uma metáfora — primeiro Deus criou a luz, que representa o Big Bang...”

Este é o ponto mais fraco para atacar o próprio judaísmo? Leia um pouco mais na Torá e você encontrará Deus matando os filhos primogênitos do sexo masculino do Egito para convencer um faraó não eleito a libertar escravos que logicamente poderiam ter sido teletransportados para fora do país. Um judeu ortodoxo está certamente familiarizado com esse episódio, pois deve ler a Torá inteira na sinagoga uma vez por ano, e este evento tem um feriado importante associado a ele. O nome “Páscoa” (“*Pessach*”<sup>28</sup>) vem de Deus passando pelas casas judaicas enquanto matava todos os primogênitos do sexo masculino no Egito.

Os judeus ortodoxos modernos são, em geral, pessoas gentis e civilizadas; muito mais civilizados do que os vários editores do Antigo Testamento. Até os antigos rabinos eram mais civilizados. Há um ritual no *Seder*<sup>29</sup> em que você tira dez gotas de vinho de sua taça, uma gota para cada uma das Dez Pragas, para enfatizar o sofrimento dos egípcios. (Claro, você deveria simpatizar com o sofrimento dos egípcios, mas não tão solidário a ponto de se levantar e dizer: “Isso não está certo! É errado fazer tal coisa!”). Em contraste, os rabinos foram suficientemente mais gentis do que os compiladores do Antigo Testamento para eles verem a dureza das Pragas. Mas a ciência era mais fraca naqueles dias e, portanto, os rabinos podiam ponderar sobre os aspectos mais desagradáveis das Escrituras sem temer que isso quebrasse totalmente sua fé.

Você nem pergunta se o incidente reflete mal em Deus, então não há necessidade de deixar escapar rapidamente “Os caminhos de Deus são misteriosos!” ou “Não somos sábios o suficiente para questionar as decisões de Deus!” ou “Matar bebês é bom quando Deus faz isso!” Essa parte da questão simplesmente não é pensada.

Acredito que pessoas religiosas instruídas permanecem fiéis porque, quando questionam sua crença, atacam somente seus pontos mais fortes, nos quais se sentem mais seguras para defender. Além disso, essas áreas são onde elas encontram conforto ao [ensaiar](#) sua defesa padrão. Provavelmente é reconfortante, por exemplo, testar a defesa prescrita para “A Ciência não diz que o universo é apenas átomos sem sentido balançando por aí?”, porque isso confirma o significado do universo e como ele flui de Deus, etc. Isso é muito mais confortável de se pensar do que uma mãe egípcia analfabeta chorando sobre o berço de seu filho massacrado.

Qualquer um que, ao questionar sua fé no judaísmo, pense espontaneamente no último exemplo, está realmente questionando-a e provavelmente não permanecerá judeu por muito mais tempo. Meu ponto aqui não é apenas criticar o Judaísmo Ortodoxo. Tenho certeza de que há alguma resposta ou outra para o Massacre dos Primogênitos, e provavelmente uma dúzia delas. Meu ponto é que, quando se trata de questionamento espontâneo, é muito mais provável que alguém ataque espontaneamente os pontos fortes com respostas reconfortantes para ensaiar, do que autoatacar espontaneamente os pontos mais fracos e vulneráveis. Da mesma forma, é provável que a pessoa pare na primeira resposta e seja consolada, em vez de continuar criticando a resposta. Um título melhor do que “Evitando os verdadeiros pontos fracos de sua crença” seria “Não pensar espontaneamente sobre as fraquezas mais dolorosas de sua crença”.

---

28 NT: Original em hebraico. *Pessach* (פסח), significa “passar por cima” ou “poupar”, referindo-se à décima praga do Egito, quando Deus “passou por cima” das casas dos israelitas. *Pessach*, também conhecido como Páscoa Judaica, é uma das principais festas do judaísmo. Comemora a libertação dos israelitas da escravidão no Egito, conforme narrado no livro do Êxodo.

29 NT: Original em hebraico. *Seder* (סדר), significa “ordem” e refere-se à refeição ritual realizada na primeira noite (ou nas duas primeiras noites fora de Israel) de *Pessach*. O *Seder* segue uma sequência específica de leituras, rituais e alimentos simbólicos que recontam a história do Êxodo e transmitem seus ensinamentos

Mais do que qualquer coisa, a força da religião é sustentada pelas pessoas que simplesmente não pensam sobre os pontos fracos reais de sua religião. As pessoas não pensam sobre os pontos fracos reais de suas crenças pelo mesmo motivo que não tocam nas bocas quentes de um fogão aceso; é doloroso.

Para fazer melhor: quando estiver duvidando de uma de suas crenças mais queridas, feche os olhos, esvazie a mente, cerre os dentes e pense deliberadamente no que mais dói. Não ensaie objeções padrão cujos contra-ataques padrão fariam você se sentir melhor. Pergunte a si mesmo o que as pessoas inteligentes que discordam diriam à sua primeira resposta e à sua segunda resposta. Sempre que você se surpreender fugindo de uma objeção em que pensou fugazmente, arraste-a para o primeiro plano de sua mente. Soque-se no plexo solar. Enfie uma faca no coração e mexa para alargar o buraco. Diante da dor, ensaie apenas isso:

O que é verdadeiro já é assim.

Reconhecê-lo não o torna pior.

Não ser honesto não o faz desaparecer.

E, porque é verdadeiro, é o que está lá para interagirmos.

Nada que seja falso está lá para ser vivido.

As pessoas conseguem suportar o que é verdadeiro,

Pois já o estão suportando<sup>30</sup>.

—Eugene Gendlin [\[1\]](#)

(Um aceno com o chapéu para Stephen Omohundro.)

## Referências

[1] Eugene T. Gendlin, *Focusing* (Bantam Books, 1982).

---

30 NT: Original em inglês.

*What is true is already so.*

*Owning up to it doesn't make it worse.*

*Not being open about it doesn't make it go away.*

*And because it's true, it is what is there to be interacted with.*

*Anything untrue isn't there to be lived.*

*People can stand what is true,*

*for they are already enduring it.*



## 75 - Parada motivada e continuação motivada



Embora eu discorde de algumas opiniões do grupo [Fast and Frugal \(Rápido e frugal\)](#) — em minha opinião, eles transformam limões em limonada com muita frequência —, também me parece que eles tendem a desenvolver os modelos mais psicologicamente realistas entre as escolas de teoria da decisão. Ao contrário de muitos experimentos que apresentam opções aos participantes, e eles simplesmente escolhem uma delas, os adeptos do Fast and Frugal reconhecem que, na realidade, você precisa gerar suas próprias opções e estudaram como as pessoas fazem isso.

Da mesma forma, muitos experimentos fornecem evidências prontas para os participantes, mas na vida real você precisa coletar evidências, o que pode ser custoso, e em algum momento precisa decidir se possui evidências suficientes para decidir. Por exemplo, ao comprar uma casa, você não é apresentado a um conjunto pré-determinado de dez casas e não é levado em uma visita guiada por todas elas antes de decidir. Em vez disso, você olha para diferentes casas, as compara, ajusta suas aspirações, considera fatores como localização e preço, decide qual casa visitar em seguida e, em algum momento, decide que já viu casas suficientes e faz uma escolha.

A distinção entre ceticismo motivado e credulidade motivada, destacada por Gilovich, mostra que as pessoas tendem a aplicar padrões mais elevados de evidência para aceitar conclusões nas quais não desejam acreditar, enquanto aplicam padrões mais baixos de evidência para aceitar conclusões nas quais desejam acreditar. Um cético motivado questiona se a evidência o compele a aceitar uma conclusão, enquanto um crente motivado questiona se a evidência lhe permite aceitar a conclusão.

Eu sugiro que existe um viés análogo na busca por uma abordagem psicologicamente realista, que é o viés da parada motivada e da continuação motivada: quando temos um motivo oculto para escolher a opção que consideramos “melhor” no momento, também temos um motivo oculto para parar, fazer uma escolha e rejeitar a consideração de outras opções. Por outro lado, quando temos um motivo oculto para rejeitar a opção que consideramos “melhor” no momento, temos um motivo oculto para suspender o julgamento, esperar por mais evidências, gerar mais opções ou encontrar algo para fazer, evitando chegar a uma conclusão.

Um exemplo histórico notável disso ocorreu no campo das estatísticas com R.A. Fisher, um renomado fundador desse campo, que afirmou que nenhuma relação causal havia sido estabelecida entre fumar e câncer de pulmão. Ele testemunhou perante o Congresso, afirmando que “correlação não implica causalidade”. Fisher sugeriu que talvez os fumantes tivessem um gene que os predisponha tanto a fumar quanto ao câncer de pulmão.

Ou talvez o emprego de Fisher como consultor de empresas de tabaco tenha lhe dado um motivo oculto para decidir que as evidências já reunidas eram insuficientes para chegar a uma conclusão e que era melhor continuar procurando. Fisher também era fumante e morreu de câncer de cólon em 1962.

(Nota *ad hominem*: Fisher era um frequentista. [Bayesianos](#) são mais razoáveis em inferir causalidade provável.)

Assim como muitas outras formas de ceticismo motivado, a continuação motivada pode tentar se disfarçar de racionalidade virtuosa. Quem poderia argumentar contra a [coleta de mais evidências?](#) Eu posso argumentar. As evidências costumam ser caras e, pior ainda, demoradas, e certamente não há nada de virtuoso em recusar-se a integrar as evidências que você já possui. [Sempre é possível mudar de ideia posteriormente.](#) (A aparente contradição pode ser resolvida da seguinte forma: passar uma hora discutindo o problema, com

a mente cuidadosamente livre de todas as conclusões, é diferente de esperar dez anos por outro estudo de US\$20 milhões.)

Quanto à parada motivada, ela ocorre em qualquer situação em que uma [terceira alternativa](#) seja temida, e também sempre que você possui um argumento com um [contra-argumento óbvio](#) que preferiria não reconhecer, bem como em outros contextos. Aparece quando você segue um curso de ação que lhe [proporciona uma sensação de bem-estar simplesmente por agir](#) e, portanto, prefere não investigar o quão bem seu plano realmente funcionou, com medo de dissipar o caloroso brilho de satisfação moral pelo qual você pagou. Ela aparece sempre que suas crenças e expectativas estão em desacordo, e então você tem motivo para temer qualquer nova evidência coletada.

A moral da história é que a decisão de encerrar um processo de busca (temporária ou permanentemente) é, assim como o próprio processo, suscetível a preconceitos e motivos ocultos. Você deve desconfiar de uma parada motivada quando encerra a pesquisa após chegar a uma conclusão confortável, mas ainda há muitas evidências rápidas e acessíveis que você não reuniu — existem sites que você pode visitar, existem contra-contra-argumentos que você pode considerar, ou você não fechou os olhos por cinco minutos, olhando o relógio, tentando pensar em uma opção melhor.

Você deve desconfiar de uma continuação motivada quando alguma evidência está se inclinando para um resultado que você não gosta, mas decide que mais evidências são necessárias — evidências dispendiosas que você sabe que não poderá reunir tão prontamente, ao contrário de algo que você pode pesquisar no Google em trinta minutos — antes de tomar uma ação desconfortável.

## 76 - Justificativas falsas



Muitos cristãos que deixaram de acreditar agora insistem em reverenciar a Bíblia como fonte de conselhos éticos. A resposta padrão dos ateus, dada por Sam Harris, é: “Você e eu sabemos que poderíamos escrever um livro com uma moralidade mais coerente e compassiva do que a Bíblia em apenas cinco minutos.” Da mesma forma, há quem tente argumentar que a Bíblia tem valor como obra literária. Mas, por que não reverenciar “O Senhor dos Anéis”, uma obra literária muito superior? Apesar das críticas à moralidade de Tolkien, “O Senhor dos Anéis” é, no mínimo, superior à Bíblia como fonte de ética. Então, por que as pessoas não usam pequenos anéis no pescoço em vez de cruzes? Até “Harry Potter” é superior à Bíblia, tanto como obra de arte literária quanto como filosofia moral. Se eu quisesse ser realmente cruel, compararia a Bíblia à série “Kushiel” de Jacqueline Carey.

“Como você pode justificar a compra de um [laptop cravejado de joias de US\\$ 1 milhão?](#)”, você pergunta ao seu amigo, “quando tantas pessoas não têm nem mesmo um laptop?” E seu amigo responde: “Mas pense nos empregos que isso proporcionará — para o fabricante do laptop, para a agência de publicidade do fabricante — e então eles comprarão refeições e cortes de cabelo — isso estimulará a economia e, eventualmente, muitas pessoas terão seus próprios laptops.” Contudo, seria ainda mais eficiente comprar 5.000 laptops do programa “One Laptop Per Child” (Um Laptop Por Criança), criando assim empregos para os fabricantes do OLPC e distribuindo os laptops diretamente.

Já mencionei antes a falha em buscar uma [terceira alternativa](#). Mas isso não é realmente uma [parada motivada](#). Chamá-la de “parada motivada” implicaria que uma busca foi realizada em primeiro lugar.

Em “[O resultado final](#)”, observei que apenas os determinantes reais de nossas crenças podem influenciar nossa precisão no mundo real, e apenas os determinantes reais de nossas ações podem influenciar nossa eficácia em alcançar nossos objetivos. Alguém que compra um laptop de um milhão de dólares estava realmente pensando: “Uau, que brilhante!”, e essa foi a única verdadeira história causal por trás de sua decisão de comprar um laptop. Nenhuma quantidade de “justificativa” pode mudar isso, a menos que essa justificativa seja um processo genuíno de pesquisa recém-realizado que altere a conclusão. Alterar a conclusão de fato. A maioria das críticas [feitas por senso de dever](#) é mais uma inspeção simbólica do que qualquer outra coisa. São como eleições livres em um país de partido único.

Para justificar genuinamente a Bíblia como um objeto de louvor com base em sua qualidade literária, você teria que realizar uma leitura neutra de vários livros até encontrar o de mais alta qualidade literária. O renome é um critério razoável para gerar candidatos, então suponho que você poderia legitimamente acabar lendo Shakespeare, a Bíblia e “Gödel, Escher, Bach”. (Caso contrário, seria uma grande coincidência encontrar a Bíblia como candidata entre um milhão de outros livros). A verdadeira dificuldade está em realizar essa “leitura neutra”. É fácil o suficiente se você não for um cristão, mas se for...

Mas é claro que nada disso aconteceu. Nenhuma pesquisa foi realizada. Colocar a “justificativa da qualidade literária” acima da linha de fundo de “Eu ♥ a Bíblia” é uma deturpação histórica de como a linha de fundo realmente se originou, como vender leite de gato como leite de vaca. A justificativa simplesmente não corresponde à origem real da conclusão.

Se você genuinamente submete sua conclusão a uma crítica que tem o potencial de desfazê-la — se a crítica realmente tem esse poder — então isso modifica o “verdadeiro algoritmo por trás” de sua conclusão. Isso altera a complexidade de sua conclusão em relação aos mundos possíveis. No entanto, as pessoas super-

estimam, e muito, a probabilidade de [mudarem realmente de ideia](#).

Com todas essas mentes abertas por aí, você pensaria que haveria mais atualizações de crenças.

Deixe-me adivinhar: sim, você admite que originalmente decidiu comprar um laptop de um milhão de dólares pensando: “Uau, brilhante!”. Sim, você admite que essa não é uma abordagem de tomada de decisão alinhada com seus objetivos declarados. Mas, desde então, você concluiu que deveria gastar seu dinheiro de maneira a fornecer laptops para o maior número possível de pessoas desfavorecidas sem acesso a um. No entanto, você simplesmente não conseguiu encontrar uma maneira mais eficiente de fazer isso além de comprar um laptop cravejado de diamantes de um milhão de dólares — porque, afinal, você está contribuindo com dinheiro para uma loja de laptops e estimulando a economia! Nada melhor do que isso!

Meu amigo, estou muito desconfiado dessa incrível coincidência. Desconfio muito que a resposta mais adequada sob esse adorável, racional e altruísta critério X seja também a ideia que surgiu originalmente desse indefensável processo Y. Se for realmente o caso, qual é a probabilidade de ela surgir de qualquer outra forma de cognição irracional?

É improvável que você tenha empregado um raciocínio falho e não tenha cometido erros.

## 77 - É essa a sua verdadeira objeção?



Ocasionalmente, acontece de alguém se deparar com algumas das minhas crenças transumanistas, que diferem das minhas ideias relacionadas à racionalidade humana. Essas crenças incluem ideias estranhas e exóticas, como superinteligência e inteligência artificial amigável. E esse alguém as rejeita.

Se esse alguém é questionado sobre o motivo da sua objeção, não é incomum que diga: “Por que devo acreditar em qualquer coisa que Yudkowsky diz? Ele não tem doutorado!”

E ocasionalmente outra pessoa, ouvindo, diz: “Ah, você deveria fazer um doutorado, para que as pessoas o ouçam”. Ou esse conselho pode até ser dado pela mesma pessoa que expressou descrença, dizendo: “Volte quando tiver um doutorado”.

Embora existam boas e más razões para se fazer um doutorado, essa é uma das ruins.

Existem muitas razões pelas quais alguém tem uma reação adversa às teses transumanistas. A maioria são questões de reconhecimento de padrões, e não de pensamento verbal: a tese é associada à “ideia estranha e bizarra” ou “ficção científica” ou “seita do fim do mundo” ou “juventude excessivamente entusiasmada”.

Assim, imediatamente, na velocidade da percepção, a ideia é rejeitada. Se, depois, alguém perguntar “Por que não?”, isso desencadeia uma busca por justificativa. Mas essa busca não necessariamente chegará à verdadeira razão — por “verdadeira razão” não quero dizer a melhor razão que poderia ser oferecida, mas sim, [as causas que foram decisivas como fato histórico, no momento exato em que a rejeição ocorreu](#). Em vez disso, a busca por justificativa encontra o fato que soa justificável: “Este palestrante não tem doutorado”.

Mas eu também não tenho doutorado quando falo sobre a racionalidade humana, então [por que a mesma objeção não é levantada nesse caso?](#)

Mais importante ainda, se eu tivesse um doutorado, as pessoas não considerariam isso um fator decisivo para acreditar em tudo o que digo. Pelo contrário, ocorreria a mesma objeção inicial, pelos mesmos motivos; e a busca por justificativa, depois, terminaria em um ponto diferente.

Eles diriam: “Por que eu deveria acreditar em você? Você é apenas mais um doutor! Existem muitos por aí. Volte quando for bem conhecido em sua área e professor titular em uma grande universidade.”

Mas as pessoas realmente acreditam em quaisquer professores de Harvard que dizem coisas estranhas? Claro que não. (Mas se eu fosse um professor em Harvard, seria de fato mais fácil chamar a atenção da mídia. Os repórteres inicialmente não inclinados a acreditar em mim — que provavelmente também não estariam inclinados a acreditar em um doutor qualquer — ainda assim fariam reportagens sobre mim, porque seria notícia que um professor de Harvard acreditasse em uma coisa tão estranha.)

Se você está dizendo coisas que soam erradas para um novato, em vez de apenas soltar baboseiras tecnológicas que parecem mágicas sobre tranças de *quarks* lépticos em dimensões  $N + 2$ ; e se o ouvinte for um estranho, não familiarizado com você pessoalmente nem com o assunto da sua área; então suspeito que o ponto no qual a pessoa média começará realmente a dar crédito, superando sua impressão inicial, puramente devido a credenciais acadêmicas, está em algum lugar próximo ao nível do Prêmio Nobel. Se tanto. Em linhas gerais, você precisa ter um nível de credencial acadêmica que se qualifique como “além do mundano”.

Pelo que sei, foi mais ou menos isso que aconteceu com Eric Drexler. Ele apresentou sua visão da nanotecnologia e as pessoas disseram: “Onde estão os detalhes técnicos?” ou “Volte quando tiver um doutorado!” E Eric Drexler passou seis anos escrevendo detalhes técnicos e obteve seu doutorado com Marvin Minsky por fazer isso. E, *Nanosystems* (Nanosistemas) é um ótimo livro. Mas será que as mesmas pessoas que disseram: “Volte quando tiver um doutorado” realmente mudaram de ideia sobre a nanotecnologia molecular? Não que eu saiba.

Da mesma forma, tem sido uma regra geral no *Machine Intelligence Research Institute* que, independentemente do que devemos fazer para aumentar nossa credibilidade, quando realmente o fazemos, nada muda significativamente. Por exemplo, quando alguém diz: “Você realiza algum tipo de desenvolvimento de código? Não estou interessado em apoiar uma organização que não desenvolve código”, mencionamos o OpenCog, mas nada muda. Ou quando alguém aponta que Eliezer Yudkowsky não possui credenciais acadêmicas, nomeamos o Professor Ben Goertzel como Diretor de Pesquisa, mas nada muda. A única coisa que realmente parece aumentar nossa credibilidade é ter pessoas famosas associadas à organização, como Peter Thiel nos financiando ou Ray Kurzweil no Conselho.

Essa pode ser uma informação importante para jovens empresas e consultores iniciantes — o motivo que clientes em potencial apontam como razão para a rejeição pode não ser o fator determinante; portanto, é preciso ponderar cuidadosamente antes de investir grandes esforços. Se o investidor disser: “Se apenas suas vendas estivessem crescendo um pouco mais rápido!” ou se o cliente em potencial disser: “Parece bom, mas você não tem o recurso X”, isso pode não ser a verdadeira objeção. Resolver isso pode, ou não, mudar alguma coisa.

E isso também é algo a ser considerado durante desacordos. Robin Hanson e eu compartilhamos a crença de que dois racionalistas não devem [concordar em discordar](#): eles não devem ter conhecimento comum de um desacordo epistêmico, a menos que algo esteja muito errado.

Suspeito que, em geral, se dois racionalistas se dispõem a resolver um desacordo que persistiu após a primeira troca, eles devem esperar encontrar que as verdadeiras fontes do desacordo são difíceis de comunicar ou difíceis de expor. Por exemplo:

- Conhecimento científico ou matemático incomum, mas bem fundamentado;
- Longas distâncias inferenciais;
- Intuições difíceis de verbalizar, talvez decorrentes de visualizações específicas;
- Zeitgeists herdados de uma profissão (que pode ter boas razões para isso);
- Padrões perceptualmente reconhecidos pela experiência;
- Meros hábitos de pensamento;
- Compromissos emocionais em acreditar em um determinado resultado;
- Medo de que um erro do passado seja refutado;
- Autoengano profundo por orgulho ou outros benefícios pessoais.

Se o tema em questão permitisse que todas as verdadeiras objeções fossem expostas facilmente, então o desacordo provavelmente seria resolvido de forma tão rápida, que nunca teria passado do primeiro encontro.

“Esta é a minha verdadeira objeção?” é algo que ambos os discordantes certamente deveriam estar se perguntando, para tornar as coisas mais fáceis para o outro. No entanto, tentativas de psicanalisar direta e publicamente o outro podem fazer com que a conversa se deteriore rapidamente, segundo minhas observações.

Ainda assim, “Essa é a sua verdadeira objeção?” deveria ser uma questão aceitável para os discordantes perguntarem humildemente, se houver alguma maneira produtiva de abordar essa subquestão. Talvez a regra pudesse ser que você pode perguntar abertamente: “Esse motivo simples e direto é sua verdadeira objeção, ou vem da intuição-X ou da influência profissional-Y?” Enquanto as possibilidades mais embaraçosas ficam na consciência do Outro, sendo sua própria responsabilidade lidar com elas.

## 78 - Verdades Entrelaçadas, Mentiras Contagiosas



Um de seus filósofos mais antigos chegou à conclusão de que uma mente plenamente competente, a partir do estudo de um único fato ou artefato pertencente a qualquer universo dado, poderia construir ou visualizar esse universo, desde o instante de sua criação até seu fim último...<sup>31</sup>

— *First Lensman* (Primeiro Patrulheiro) [1]

Se qualquer um de vocês se concentrar em um único fato, ou pequeno objeto, como um seixo ou a semente de uma planta ou outra criatura, por um período tão curto quanto cem de seus anos, você começará a perceber sua verdade<sup>32</sup>.

— *Gray Lensman* (Patrulheiro Cinzento) [2]

Tenho razoável certeza de que um único seixo, retirado de uma praia da nossa própria Terra, não especifica os continentes e países, nem a política e as pessoas deste planeta. Outros planetas no espaço e no tempo, outros [ramos de Everett](#), gerariam o mesmo seixo. Por outro lado, a identidade de um único seixo parece incluir as nossas leis da física. Nesse sentido, a totalidade do nosso Universo—todos os ramos de Everett—seria implicada pelo seixo. (Se, como parece provável, não houver variáveis verdadeiramente livres.)

Portanto, um único seixo provavelmente não implica toda a nossa Terra. Mas um único seixo implica uma enorme quantidade de informações. A partir do estudo desse único seixo, você poderia ver as leis da física e tudo o que elas implicam. Pensando sobre essas leis da física, você pode ver que planetas se formarão, e pode deduzir que o seixo veio de um desses planetas. As formações internas de cristais e moléculas do seixo se formaram sob gravidade, o que lhe diz algo sobre a massa do planeta; a mistura de elementos no seixo lhe diz algo sobre a formação do planeta.

Não sou geólogo, então não sei quais mistérios os geólogos conhecem. Mas acho muito fácil imaginar mostrar um seixo a um geólogo e dizer: “Este seixo veio de uma praia em *Half Moon Bay* (Baía da Meia Lua)”, e o geólogo imediatamente dizer: “Estou confuso” ou até mesmo “Você está mentindo”. Talvez seja o tipo errado de rocha, ou o seixo não esteja desgastado o suficiente para ser de uma praia—eu não conheço seixos bem o suficiente para adivinhar as conexões e assinaturas pelas quais eu poderia ser pego, e esse é o ponto.

“Somente Deus pode contar uma mentira verdadeiramente plausível.” Pergunto-me se já houve uma religião que desenvolveu isso como um provérbio. Eu (de forma falseável) diria que não: é um sentimento racionalista, mesmo que você o coloque em metáfora teológica. Dizer “tudo está interconectado com todo o resto, porque Deus fez o mundo inteiro e o sustenta” pode gerar alguns sentimentos agradáveis e aconchegantes durante o sermão, mas não te leva muito longe quando se trata de atribuir seixos a praias.

---

31 NT: Texto original em inglês. *One of your very early philosophers came to the conclusion that a fully competent mind, from a study of one fact or artifact belonging to any given universe, could construct or visualize that universe, from the instant of its creation to its ultimate end...*

32 NT: Texto original em inglês. *If any one of you will concentrate upon one single fact, or small object, such as a pebble or the seed of a plant or other creature, for as short a period of time as one hundred of your years, you will begin to perceive its truth.*

Uma moeda na Terra exerce uma aceleração gravitacional na Lua de cerca de  $4,5 \times 10^{-31} \text{ m/s}^2$ , então, em certo sentido, não é muito errado dizer que cada evento está entrelaçado com todo o seu cone de luz passado. E como as inferências podem se propagar para trás e para frente através de redes causais, os entrelaçamentos epistêmicos podem facilmente cruzar as fronteiras dos cones de luz. Mas eu não gostaria de ser o [astrônomo forense](#) que tivesse que olhar para a Lua e descobrir se a moeda caiu cara ou coroa—a influência é muito menor do que a incerteza quântica e o ruído térmico.

Se você dissesse “Tudo está entrelaçado com alguma outra coisa” ou “Tudo está entrelaçado inferencialmente e alguns entrelaçamentos são muito mais fortes que outros”, você poderia ser realmente sábio em vez de apenas “[Profundamente Sábio](#)”.

Fisicamente, cada evento é, de certa forma, a soma de todo o seu cone de luz passado, sem fronteiras ou limites. Mas a lista de entrelaçamentos perceptíveis é muito mais curta, e ela te dá algo como uma rede. Essa regularidade de alto nível é o que eu chamo de Grande Teia da Causalidade.

Eu uso essas Letras Maiúsculas de forma um tanto irônica, talvez; mas se alguma coisa merece Letras Maiúsculas, certamente a Grande Teia da Causalidade está na lista.

“Oh, que teia emaranhada tecemos, quando primeiro praticamos para enganar”, disse Sir Walter Scott. Nem todas as mentiras saem do controle—não vivemos em um universo tão justo. Mas acontece ocasionalmente de alguém mentir sobre um fato, e então ter que mentir sobre um fato entrelaçado, e depois sobre outro fato entrelaçado com aquele:

“Onde você estava?”

“Ah, eu estava em uma viagem de negócios.”

“Sobre o que era a viagem de negócios?”

“Não posso te dizer; são negociações confidenciais com um cliente importante.”

“Oh—eles estão te incluindo nessas? Boas notícias! Devo ligar para seu chefe para agradecer-lo por te adicionar.”

“Desculpe—ele não está no escritório agora...”

Os seres humanos, que não são deuses, muitas vezes falham em imaginar todos os fatos que precisariam distorcer para contar uma mentira verdadeiramente plausível. “[Deus me engravidou](#)” soava um pouco mais provável nos velhos tempos, antes que nossos modelos do mundo contivessem (citações de) cromossomos Y. Muitas mentiras semelhantes, hoje, podem explodir quando os testes genéticos se tornarem mais comuns. Estupradores foram condenados, e falsos acusadores expostos, anos depois, com base em evidências que não perceberam que poderiam deixar. Um estudante de biologia evolutiva pode ver a assinatura do design da seleção natural em cada lobo que persegue um coelho; e cada coelho que foge; e cada abelha que pica em vez de transmitir um aviso educado—mas as fraudes dos criacionistas soam plausíveis para eles, tenho certeza.

Nem todas as mentiras são descobertas, nem todos os mentirosos são punidos; não vivemos em um universo tão justo. Mas nem todas as mentiras são tão seguras quanto seus mentirosos acreditam. Quantos pecados se tornariam conhecidos por uma superinteligência bayesiana, eu me pergunto, se ela fizesse uma varredura nanotecnológica (não destrutiva?) da Terra? No mínimo, todas as mentiras das quais ainda existe alguma evidência em qualquer cérebro. Algumas dessas mentiras podem se tornar conhecidas mais cedo do que isso, se os neurocientistas algum dia conseguirem construir um detector de mentiras realmente bom via neuroimagem. Paul Ekman (um pioneiro no estudo de pequenos movimentos musculares faciais) provavelmente poderia ler uma fração considerável das mentiras do mundo agora mesmo, se tivesse uma chance.

Nem todas as mentiras são descobertas, nem todos os mentirosos são punidos. Mas a Grande Teia é muito comumente subestimada. Apenas o conhecimento que os humanos já acumularam [levaria muitas vidas humanas para aprender](#). Qualquer um que pense que um não-Deus pode contar uma mentira perfeita, sem riscos, está subestimando o emaranhamento da Grande Teia.



A honestidade é a melhor política? Não sei se eu iria tão longe: Mesmo na minha ética, às vezes é aceitável ficar calado. Mas comparado a mentiras descaradas, tanto a honestidade quanto o silêncio envolvem menos exposição a riscos recursivamente propagantes que você não sabe que está correndo.

## **Referências**

- [1] Edward Elmer Smith and A. J. Donnell, *First Lensman* (Old Earth Books, 1997).
- [2] Edward Elmer Smith and Ric Binkley, *Gray Lensman* (Old Earth Books, 1998).

## 79 - Sobre mentiras e explosões de Cisne Negro



O juiz Marcus Einfeld, de 70 anos, era Conselheiro da Rainha desde 1977, foi nomeado Tesouro Vivo da Austrália em 1997 e recebeu o Prêmio da Paz das Nações Unidas em 2002. Ele foi presidente fundador da Comissão de Direitos Humanos e Igualdade de Oportunidades da Austrália. Aposentou-se há alguns anos, mas é frequentemente convocado para julgar casos importantes...

... [Foi condenado](#) a dois anos de prisão por uma série de perjúrios e mentiras, que [tiveram início com uma multa de 36 libras esterlinas por exceder a velocidade em 10 km/h.](#)

Essa teoria, aparentemente virtuosa, sobre pessoas honestas não serem boas em mentir, e sobre vestígios entrelaçados deixados em algum lugar, resultando em uma falha épica do tipo Cisne Negro, de fato tem alguns exemplos na vida real. No entanto, é evidente que há uma seleção tendenciosa nos casos que chegam ao nosso conhecimento.

## 80 - Epistemologia do lado negro



Se uma vez você contar uma mentira, a verdade se tornará seu inimigo.

[Já falei antes](#) sobre a ideia de que, sendo a verdade interligada, as mentiras são contagiosas. Se você pegar uma pedra na entrada da sua garagem e disser a um geólogo que a encontrou em uma praia você sabe o que um geólogo sabe sobre rochas? Eu não sei. No entanto, posso suspeitar que uma pedra desgastada pela água não se pareceria com uma gota de lava congelada de uma erupção vulcânica. Você sabe de onde realmente veio a pedra na entrada da sua casa? As coisas carregam marcas de seus lugares em um universo ordenado; dentro dessa teia, uma mentira está fora de lugar. (Na verdade, [um geólogo nos comentários](#) menciona que a maioria das pedras nas calçadas são retiradas das praias, então eles não poderiam dizer a diferença entre uma pedra na entrada da garagem e uma pedra na praia, mas poderiam diferenciar entre uma pedra da montanha e um seixo da praia. Caso em questão...)

O que pode parecer uma verdade arbitrária para uma mente — algo que poderia ser facilmente substituído por uma mentira plausível — pode ser sustentado por uma série de conexões aos olhos de um conhecimento mais amplo. Para um criacionista, a ideia de que a vida foi moldada por “design inteligente” em vez de “seleção natural” pode soar como uma preferência pessoal. Já para um biólogo, argumentar plausivelmente que um organismo foi projetado de forma inteligente exigiria mentir sobre quase todos os aspectos desse organismo. Para argumentar plausivelmente que os “humanos” foram projetados de forma inteligente, seria necessário mentir sobre o design da retina humana, a arquitetura do cérebro humano e as proteínas unidas por fracas forças de van der Waals em vez de fortes ligações covalentes...

Ou você poderia simplesmente mentir sobre a teoria da evolução, o qual é o caminho seguido pela maioria dos criacionistas. Em vez de mentir sobre os nós conectados na rede, eles mentem sobre as leis gerais que governam essas conexões.

E para encobrir isso, eles mentem sobre as regras da ciência — como o que significa chamar algo de “teoria” ou o que significa para um cientista dizer que não tem certeza absoluta.

Assim, eles passam de, mentir sobre fatos específicos, para mentir sobre leis gerais, e depois para mentir sobre as regras do raciocínio. Para mentir sobre se os humanos evoluíram, você deve mentir sobre a evolução; e então você tem que mentir sobre as regras científicas que limitam nossa compreensão da evolução.

Mas como poderia ser diferente?

Assim como um ser humano estaria fora do lugar em uma comunidade de formas de vida genuinamente projetadas de forma inteligente, e seria necessário mentir sobre as regras da evolução para fazer isso parecer o oposto, as próprias crenças sobre o criacionismo estão descontextualizadas na ciência.

Essas crenças não seriam encontradas em uma mente bem ordenada, assim como palmeiras não crescem em geleiras. Portanto, seria necessário romper as barreiras que as proibiriam.

O que nos leva ao caso do autoengano.

Uma única mentira que você conta a si mesmo pode parecer bastante plausível quando não se conhece as regras que governam o pensamento, ou sequer se sabe que existem regras; e a [escolha](#) parece tão arbitrária quanto escolher um sabor de sorvete, tão isolada quanto uma pedra na praia...

... daí alguém questiona sua crença, usando as regras de raciocínio que aprendeu. Eles perguntam: “Onde estão as suas evidências?”

E você responde: “O quê? Por que eu preciso de evidências?”

Então eles dizem: “Em geral, as crenças exigem evidências.”

Este argumento, claramente, é um [soldado lutando do lado inimigo](#), que você deve derrotar. Então, você diz: “Eu discordo! Nem todas as crenças exigem evidências. Especificamente, crenças sobre dragões não precisam de evidências. Quando se trata de dragões, você pode acreditar no que quiser. Portanto, não preciso de evidências para acreditar que há um dragão na minha garagem.”

E o outro diz: “Ei! Você não pode simplesmente excluir dragões assim. Há uma razão para a regra de que as crenças exigem evidências. Para criar um mapa preciso da cidade, você deve percorrer suas ruas e desenhar linhas no papel que correspondam ao que você vê. Isso não é uma exigência legal arbitrária — se você simplesmente se sentar em sua sala de estar e desenhar linhas aleatórias no papel, o mapa estará errado, com uma probabilidade extremamente alta. Isso se aplica tanto ao mapa ou ao dragão quanto a qualquer outra coisa.”

Então agora, essa explicação do porquê as crenças exigem evidências também é um soldado inimigo. Você diz: “Errado com uma probabilidade extremamente alta? [Então ainda há uma chance, certo?](#) Não preciso acreditar se não for uma [certeza absoluta](#).”

Ou talvez você até comece a suspeitar, você mesmo, que “crenças exigem evidências”. Mas isso ameaça uma mentira que você valoriza; então você rejeita o amanhecer interior, empurrando o Sol de volta para baixo do horizonte.

Ou talvez você já tenha ouvido o provérbio “crenças exigem evidências” e achou sábio o suficiente para endossá-lo publicamente. Mas nunca ocorreu a você, até que outra pessoa chamou sua atenção, que esse provérbio poderia se aplicar à sua crença de que há um dragão em sua garagem. Então você pensa rapidamente e diz: “O dragão está em um magistério separado”.

Ter falsas crenças não é nada bom, mas não precisa ser permanentemente incapacitante — se, ao descobrir seu erro, você o superar. O perigo reside em ter uma crença falsa que você acredita que deve ser protegida como uma crença — uma crença na crença, acompanhada ou não de uma crença real.

Uma única Mentira Que Precisa Ser Protegida pode bloquear o progresso de alguém em direção à racionalidade avançada. Não, não é uma diversão inofensiva.

Assim como o próprio mundo é muito mais emaranhado do que parece à superfície, da mesma forma existem regras de raciocínio mais rigorosas, restringindo as crenças mais fortemente, do que os inexperientes suspeitam. O mundo é intrinsecamente entrelaçado, governado por leis gerais, assim como as crenças racionais.

Pense no que seria necessário para negar a evolução ou o heliocentrismo — todas as verdades interconectadas e leis governantes que você não teria permissão para conhecer. Então você pode imaginar como um único ato de autoengano pode bloquear todo o meta-nível da busca pela verdade, uma vez que sua mente começa a se sentir ameaçada ao perceber as conexões. Proibindo todos os níveis intermediários e superiores da Arte do racionalista. Criando, em seu lugar, um vasto complexo de anti-leis, regras anti-pensamento e justificativas gerais para acreditar no que não é verdade.

Steven Kaas disse: “Promover crenças menos que maximamente precisas é um ato de sabotagem. Não faça isso com ninguém, a menos que você também furasse os pneus deles.” Dar a alguém uma crença falsa para proteger — convencê-lo de que a crença em si deve ser defendida de qualquer pensamento que pareça ameaçá-la — bem, você não deve fazer isso com ninguém, a menos que também fizesse uma lobotomia frontal neles.

Uma vez que você conta uma mentira, a verdade se torna sua inimiga; e toda verdade conectada a essa verdade, e todos os aliados da verdade em geral; todos eles você deve combater para proteger a mentira\*\*, seja\*\* você mentindo para os outros ou para si mesmo.

Você tem que negar que as crenças exigem evidências\*\*, e então\*\* você tem que negar que os mapas devem refletir os territórios, e depois você tem que negar que a verdade é algo bom...

Assim, surge o Lado Negro.

Eu me preocupo de que as pessoas não estejam cientes disso ou não estejam suficientemente cautelosas — que enquanto vagamos por nosso mundo humano, podemos esperar encontrar uma epistemologia sistematicamente ruim.

Os memes de “como pensar” que circulam pela internet, os [pensamentos armazenados](#) da [Sabedoria Profunda](#) — alguns deles são conselhos valiosos elaborados por racionalistas. Mas outras noções foram inventadas para proteger uma mentira ou autoengano: geradas pelo Lado Negro.

“Todo mundo tem direito à sua própria opinião.” Quando você pensa sobre isso, de onde surgiu esse provérbio? É algo que alguém diria no curso de proteger uma verdade, ou no curso de se proteger da verdade? Mas as pessoas não se animam e dizem: “Aha! Sinto a presença do Lado Negro!” Até onde posso perceber, não é amplamente percebido que o Lado Negro está por aí.

Mas como poderia ser diferente? Seja enganando os outros, ou apenas a si mesmo, a Mentira Que Precisa Ser Protegida se espalhará recursivamente pela rede de causalidade empírica, e pela rede de regras empíricas gerais, e pelas próprias regras do raciocínio, e pelo entendimento subjacente a essas regras. Se houver uma boa epistemologia no mundo, e também mentiras ou autoengano que as pessoas estão tentando proteger, então surgirá uma má epistemologia para combater a boa. Dificilmente poderíamos esperar, neste mundo, encontrar o Lado da Luz sem o Lado Negro; há o Sol, e aquilo que se encolhe e gera uma Sombra que o encobre.

Note que essas [pessoas não são necessariamente más](#). A grande maioria dos que repetem a “Sabedoria Profunda”, estão mais enganados do que enganadores, mais autoenganados do que enganadores. Eu acho.

E certamente não é minha intenção fornecer um [Contra-argumento Totalmente Geral](#), de modo que toda vez que alguém lhe oferecer uma epistemologia que você não goste, você diga: “Ah, alguém do Lado Negro inventou isso.” É uma das regras do Lado da Luz que você deve refutar a proposição por si mesma, não acusando seu inventor de [más intenções](#).

Mas o Lado Negro está por aí. O medo é o caminho que leva a ele, e uma traição pode transformá-lo. Nem todos que usam túnicas são *Jedi* ou impostores; também existem os *Lordes Sith*, mestres e aprendizes involuntários. Esteja avisado, esteja atento.

Quanto a listar memes comuns gerados pelo Lado Negro — não crenças falsas aleatórias, me entenda bem, mas sim má epistemologia, as Defesas Genéricas do Fracasso — bem, vocês gostariam de tentar, queridos leitores?



**H - Contra o duplipensar**



## 81 - Pensamento único



Lembro-me do momento exato em que iniciei minha jornada como racionalista.

Não foi durante a leitura de *Surely You're Joking, Mr. Feynman* (Com certeza você está brincando, Sr. Feynman) ou qualquer outro trabalho existente sobre racionalidade; estes eu simplesmente aceitei como óbvios. A jornada começa quando você percebe uma grande falha em sua arte atual e descobre um impulso para melhorar, para desenvolver novas habilidades além das úteis, mas inadequadas, encontradas nos livros.

Nos últimos momentos da minha primeira vida, eu tinha quinze anos e repassava uma lembrança prazerosamente hipócrita de uma época em que eu era muito mais jovem. Minhas memórias daquele tempo são vagas; tenho uma imagem mental, mas não me recordo exatamente que idade eu tinha. Acredito que teria seis ou sete anos e que o evento original ocorreu durante um acampamento de verão.

O que aconteceu naquela ocasião foi que um monitor do acampamento, um adolescente do sexo masculino, fez com que nós, meninos muito mais novos, formássemos uma fila e propôs o seguinte jogo: o garoto no final da fila rastejaria entre nossas pernas e nós o bateríamos enquanto ele passava, e então seria a vez do próximo garoto de oito anos no final da fila. (Talvez seja apenas porque perdi minha inocência juvenil, mas não posso deixar de me perguntar...) Recusei participar desse jogo e me mandaram sentar no canto.

Essa lembrança - de recusar bater e de apanhar - passou a simbolizar para mim que, mesmo nessa idade muito precoce, eu me negava a encontrar prazer em machucar os outros. Que eu não compraria uma palmada na bunda de outra pessoa ao preço de uma palmada na minha própria; não pagaria com dor a oportunidade de infligir dor. Eu me recusei a jogar um jogo de soma negativa.

E então, aos quinze anos, de repente percebi que isso não era verdade. Eu não havia recusado por uma posição de princípios contra jogos de soma negativa. Eu descobri o Dilema do Prisioneiro muito cedo na vida, mas não aos sete anos. Eu recusei participar simplesmente porque não queria me machucar, e ficar sentado no canto era um preço aceitável a pagar por não me machucar.

Mais importante, percebi que sempre soube disso - que a verdadeira lembrança sempre esteve escondida em um canto da minha mente, meu olho mental lançando um olhar para ela por uma fração de segundo e depois desviando o olhar.

No meu primeiro passo ao longo do Caminho, captei a sensação - generalizada sobre a experiência subjetiva - e disse: "Então é assim que se sente empurrar uma verdade indesejada para um canto da minha mente! Agora vou notar cada vez que eu fizer isso e vou limpar todos os meus cantos!"

Chamei essa disciplina de "pensamento único", inspirado no duplipensar de Orwell. No [duplipensar](#), você esquece e, em seguida, esquece que esqueceu. No pensamento único, você percebe que está esquecendo e então se lembra. Você mantém apenas um único pensamento não contraditório em sua mente de cada vez.

O "pensamento único" foi a primeira nova habilidade racionalista que criei, sobre a qual não havia lido nos livros. Duvido que seja original no sentido de prioridade acadêmica, mas felizmente isso não é necessário.

Ah, e o meu eu de quinze anos gostava de nomear as coisas.

As profundezas aterrorizantes do viés de confirmação continuam e continuam. Não eternamente, já que o cérebro é de complexidade finita, mas por tempo suficiente para parecer uma eternidade. Você continua descobrindo (ou lendo sobre) novos mecanismos pelos quais seu cérebro empurra as coisas para fora do caminho.

Mas o meu eu jovem varreu muitos cantos com aquela primeira vassoura.



## 82 - Duplipensar (escolhendo ser tendencioso)



Um pedaço oblongo de papel jornal apareceu entre os dedos de O'Brien. Por cerca de cinco segundos, permaneceu dentro do campo de visão de Winston. Era uma fotografia e não havia dúvidas sobre sua identidade. Era a fotografia. Era outra cópia da fotografia de Jones, Aaronson e Rutherford na festa em Nova York, que ele encontrou há onze anos e destruiu imediatamente. Por apenas um instante, ela estava diante de seus olhos e, em seguida, desapareceu de vista novamente. Mas ele havia visto, sem dúvida alguma. Ele fez um esforço desesperado e agonizante para libertar a metade superior de seu corpo. Era impossível mover-se um centímetro em qualquer direção. No momento, ele até esqueceu o mostrador. Tudo o que ele queria era segurar a fotografia em seus dedos novamente, ou pelo menos vê-la.

"Isso existe!" ele exclamou.

"Não", disse O'Brien.

Ele atravessou a sala. Havia um buraco vazio na parede oposta. O'Brien levantou a grade. Invisível, o frágil pedaço de papel rodopiou na corrente de ar quente; estava desaparecendo em um lampejo de chama. O'Brien afastou-se da parede.

"Cinzas", disse ele. "Nem mesmo cinzas reconhecíveis. Pó. Isso não existe. Nunca existiu."

"Mas existiu! Ele existe! Está na memória. Eu me lembro. Você se lembra disso."

"Não me lembro", disse O'Brien.

O coração de Winston afundou. Aquilo era o duplipensar. Ele sentiu uma impotência mortal. Se ele pudesse ter certeza de que O'Brien estava mentindo, talvez isso não importasse. Mas era perfeitamente possível que O'Brien tivesse realmente esquecido a fotografia. E se fosse esse o caso, então ele também teria esquecido sua negação de se lembrar e até mesmo o ato de esquecer. Como alguém poderia ter certeza de que não era apenas um truque? Talvez aquela distorção insana na mente pudesse realmente ocorrer, esse foi o pensamento que o derrotou.

—George Orwell, 1984<sup>33</sup> [1]

---

33 NT.: Texto original em inglês. *An oblong slip of newspaper had appeared between O'Brien's fingers. For perhaps five seconds it was within the angle of Winston's vision. It was a photograph, and there was no question of its identity. It was the photograph. It was another copy of the photograph of Jones, Aaronson, and Rutherford at the party function in New York, which he had chanced upon eleven years ago and promptly destroyed. For only an instant it was before his eyes, then it was out of sight again. But he had seen it, unquestionably he had seen it! He made a desperate, agonizing effort to wrench the top half of his body free. It was impossible to move so much as a centimetre in any direction. For the moment he had even forgotten the dial. All he wanted was to hold the photograph in his fingers again, or at least to see it. "It exists!" he cried. "No," said O'Brien. He stepped across the room. There was a memory hole in the opposite wall. O'Brien lifted the grating. Unseen, the frail slip of paper was whirling away on the current of warm air; it was vanishing in a flash of flame. O'Brien turned away from the wall. "Ashes," he said. "Not even identifiable ashes. Dust. It does not exist. It never existed." "But it did exist! It does exist! It exists in memory. I remember it. You remember it." "I do not remember it," said O'Brien. Winston's heart sank. That was doublethink. He had a feeling of deadly helplessness. If he could have been certain that O'Brien was lying, it would not have seemed to matter. But it was perfectly possible that O'Brien had really forgotten the photograph. And if so, then already he would have*

E se o autoengano nos ajudar a ser felizes? E se simplesmente ignorar e superar o preconceito nos tornar - pasme! - infelizes? Certamente, a verdadeira sabedoria seria a racionalidade de segunda ordem, escolhendo quando ser racional. Dessa forma, você pode decidir quais vieses cognitivos devem governá-lo, a fim de maximizar sua felicidade.

Deixando a moralidade de lado, duvido que seja realmente possível ter um deslocamento mental tão extravagante.

A racionalidade de segunda ordem implica que, em algum momento, você pensará consigo mesmo: “Acreditarei irracionalmente que ganharei na loteria para me fazer feliz”. Mas não temos controle direto sobre nossas crenças. Não é possível acreditar que o céu é verde por um ato de vontade. Talvez você consiga acreditar que acreditou — embora eu tenha dificultado isso para você, apontando a diferença. (De nada!) Talvez você até possa acreditar que foi feliz e se enganou; mas, na verdade, você não estaria feliz nem se autoenganaria.

Para que a racionalidade de segunda ordem seja verdadeiramente racional, você precisa primeiro ter um bom modelo da realidade para extrapolar as consequências da racionalidade e da irracionalidade. Se você escolher ser irracional em primeira instância, precisará abandonar essa visão precisa. E então esquecer o ato de esquecer. Não planejo cometer a falácia lógica de generalizar a partir de evidências fictícias, mas acho que Orwell fez um bom trabalho em antecipar onde esse caminho leva.

Você não pode conhecer as consequências de ser tendencioso até que já esteja envolvido. E então é tarde demais para o autoengano.

A alternativa é escolher permanecer tendencioso cegamente, sem uma ideia clara das consequências. Isso não é racionalidade de segunda ordem. É uma estupidez deliberada.

Seja irracionalmente otimista sobre suas habilidades de direção e você ficará alegremente despreocupado enquanto os outros suam e temem. Você não precisará lidar com a inconveniência do cinto de segurança. Você ficará feliz e despreocupado por um dia, uma semana, um ano. Depois, sofra um acidente e passe o resto da vida desejando poder coçar uma coceira em um membro fantasma. Ou fique paralisado do pescoço para baixo. Ou, morra. Não é inevitável, mas é possível; quão provável é isso? Você não pode fazer essa troca racionalmente a menos que conheça suas verdadeiras habilidades de direção para avaliar em que perigo está se colocando. Você não pode fazer essa troca racionalmente a menos que esteja ciente de vieses como [negligenciar a probabilidade](#).

Não importa quantos dias se passem em feliz ignorância, um único erro é suficiente para arruinar uma vida humana e compensar cada centavo que você ganhou nos trilhos da estupidez.

Um dos principais conselhos que dou aos aspirantes a racionalistas é: “Não tente ser esperto” e “Ouça aquelas dúvidas silenciosas e incômodas”. Se você não sabe, não sabe o que não sabe, não sabe o quanto não sabe e não sabe o quanto precisava saber.

Não há racionalidade de segunda ordem. Existe apenas um salto cego para o que pode ou não ser um poço de lava ardente. Quando você finalmente souber, será tarde demais para a cegueira.

No entanto, as pessoas negligenciam isso porque não sabem o que não sabem. Os desconhecidos “desconhecidos” não estão disponíveis. Elas não se concentram na área em branco do mapa, mas a tratam como se fosse um território vazio. Quando consideram pular no escuro, consultam sua memória em busca de perigos e não encontram poços de lava ardente no mapa em branco. Por que não pular?

Eu já estive lá. Já tentei isso. Me queimei. Não tente ser esperto.

Certa vez, compartilhei com uma amiga que suspeitava que a felicidade da estupidez fosse superestimada. Ela balançou a cabeça seriamente e disse: “Não, não é; realmente não é.”

Pode ser que existam pessoas estúpidas e felizes por aí. Talvez elas sejam mais felizes do que você.

---

*forgotten his denial of remembering it, and forgotten the act of forgetting. How could one be sure that it was simple trickery? Perhaps that lunatic dislocation in the mind could really happen: that was the thought that defeated him.*

A vida não é justa, e você não ficará mais feliz alimentando-se de inveja do que não pode ter. Suspeito que a grande maioria dos leitores do *Overcoming Bias* (Superando o viés) não conseguiria alcançar a “felicidade da estupidez” mesmo se tentasse. Esse caminho está fechado para você. Você nunca poderá atingir esse nível de ignorância, não pode esquecer o que sabe, não pode deixar de ver o que vê.

A felicidade da estupidez está fora do seu alcance. Você nunca a experimentará sem um dano cerebral real, e talvez nem mesmo assim. Talvez você deva se perguntar se a felicidade da estupidez é a suprema felicidade que um ser humano pode almejar, mas isso não importa. Esse caminho está fechado para você, se é que algum dia esteve aberto.

Agora, tudo o que lhe resta é aspirar à felicidade que um racionalista pode alcançar. Acredito que ela possa ser maior, no final das contas. Existem caminhos limitados e caminhos abertos; planícies para descansar e montanhas para escalar. E mesmo que escalar exija mais esforço, a montanha se eleva mais alto no final.

Também há mais na vida do que a felicidade, e suas decisões podem envolver outras formas de felicidade além da sua.

Mas isso é discutível. No momento em que você percebe que tem uma escolha, essa escolha desaparece. Você não pode deixar de ver o que vê. O outro caminho está fechado.

## Referências

[1] George Orwell, 1984 (Signet Classic, 1950).

## 83 - Não, realmente, eu me autoenganei



Recentemente, tive uma conversa com uma pessoa intrigante, que é difícil de descrever em termos simples. Formalmente, ela era uma judia ortodoxa, mas havia algo a mais nela que chamava a atenção. Era altamente inteligente e tinha conhecimento de algumas evidências arqueológicas que contradiziam sua religião, bem como dos argumentos superficiais que as pessoas religiosas geralmente conhecem. Por exemplo, ela sabia que os nomes Mordecai, Ester, Hamã e Vasti não constavam nos registros históricos persas, mas estava ciente de uma antiga lenda persa correspondente que envolvia os deuses babilônicos Marduque e Ishtar, e os deuses elamitas rivais Humman e Vasti. Apesar desse conhecimento, ela ainda celebrava o *Purim*<sup>34</sup>. Essa mulher era uma daquelas pessoas religiosas altamente inteligentes que vivem em um estado de contradição interna por anos, elaborando e refinando suas crenças até que o interior de suas mentes se assemelhe a uma pintura de M.C. Escher.

A maioria das pessoas nessa situação fingiria ser demasiado sábia ou ocupada para conversar com um ateu, mas ela estava disposta a dedicar algumas horas a uma discussão sincera. Como resultado dessa interação, agora compreendo pelo menos mais um aspecto do autoengano que não havia considerado explicitamente antes: você não precisa realmente se enganar, contanto que acredite que se enganou. Podemos chamar isso de “crença no autoengano”.

Durante o ensino médio, essa mulher se considerava atea. Porém, ela tomou a decisão consciente de agir como se acreditasse em Deus. E então, conforme ela me disse com total sinceridade, ao longo do tempo, ela começou realmente a acreditar em Deus. Do meu ponto de vista, ela está completamente equivocada a esse respeito. Durante nossa conversa, ela repetidamente afirmava: “Eu acredito em Deus”, mas nunca disse: “Deus existe”. Quando perguntei a ela por que era religiosa, suas respostas sempre se concentravam nas consequências de acreditar em Deus, nunca nas consequências da existência de Deus. Ela nunca disse: “Deus vai me ajudar”, mas sim: “Minha crença em Deus me ajuda”. E quando eu argumentei que “Alguém que busca apenas a verdade e observa nosso universo não inventaria Deus sequer como uma hipótese”, ela concordou abertamente comigo.

Na verdade, ela não se enganou acreditando que Deus existe ou que a religião judaica é verdadeira, pelo menos não que eu possa perceber. No entanto, acredito que ela realmente acredita que se enganou. Portanto, embora ela não obtenha nenhum benefício prático por acreditar em Deus, ela sinceramente acredita que se autoenganou ao adotar essa crença e, conseqüentemente, espera receber os benefícios associados ao autoengano. Suponho que isso possa produzir o mesmo efeito placebo que a crença genuína em Deus.

Essa pode ser a explicação para sua motivação em defender sinceramente sua afirmação de que acreditava em Deus diante do meu ceticismo. Ela nunca disse nada como: “Ah, e, por sinal, Deus realmente existe”, nem demonstrou interesse em discutir a existência de Deus.

---

34 NT: do original hebraico, *Purim* (פּוּרִים) é uma festividade judaica que comemora a salvação dos judeus do Império Persa, conforme narrado no Livro de Ester na Bíblia Hebraica.

## 84 - Crença no autoengano



Falei sobre minha conversa com uma mulher judia nominalmente ortodoxa, que defendeu com fervor a afirmação de que acreditava em Deus, embora não parecesse ter uma crença genuína nele.

Enquanto a questionava sobre os benefícios que ela achava advirem da crença em Deus, apresentei a [Ladainha de Tarski](#) — que, na verdade, é uma família infinita de litanias, sendo um exemplo específico:

*“Se o céu é azul,  
desejo acreditar que “o céu é azul”.  
Se o céu não é azul,  
desejo acreditar que “o céu não é azul”.*

“Essa não é a minha filosofia”, ela me disse.

“Imaginei que não fosse”, respondi. “Só estou perguntando — assumindo que Deus não existe, o que não sabemos, você ainda deveria acreditar em Deus?”

Ela hesitou. Parecia que realmente estava ponderando sobre isso, o que me surpreendeu.

“Então, é uma questão contrafactual...” ela disse lentamente.

Naquela época, pensei que ela tinha dificuldade em se permitir visualizar um mundo em que Deus não existisse, devido ao seu apego a um mundo que incluía Deus.

Agora, porém, suspeito que ela enfrentava dificuldades em visualizar uma diferença entre a aparência do mundo se Deus existisse ou não, porque todos os seus pensamentos estavam centrados em sua crença em Deus, mas seu modelo causal do mundo não incluía Deus como um elemento. Portanto, ela poderia facilmente responder “Como o mundo seria diferente se eu não acreditasse em Deus?”, mas não “Como o mundo seria diferente se Deus não existisse?”

Naquela época, ela não respondeu a essa pergunta. No entanto, ela apresentou um contraexemplo para a Ladainha de Tarski:

Ela disse: “Acredito que as pessoas são melhores do que realmente são”.

Tentei explicar que se você disser: “As pessoas são más”, isso significa que você acredita que as pessoas são más, e se você disser: “Acredito que as pessoas são boas”, isso significa que você acredita que as pessoas são boas. Portanto, dizer “As pessoas são más e eu acredito que as pessoas são boas” significa que você acredita que as pessoas são más, mas acredita acreditar que as pessoas são boas.

Citei para ela:

“Se houvesse um verbo que significasse “acreditar falsamente”, ele não teria uma primeira

pessoa significativa no presente do indicativo.”<sup>35</sup>

— Ludwig Wittgenstein [1]

Ela disse, sorrindo: “Sim, acredito que as pessoas são mais amáveis do que realmente são. Apenas pensei que deveria expressar dessa maneira para você”.

“Acho que a vovó deveria dar uma boa olhada em você, Walter”, disse a babá. “Acredito que sua mente está completamente emaranhada, como uma bola de barbante que caiu<sup>36</sup>”.

— Terry Pratchett, *Maskerade* (Disfarce) [2]

E posso digitar as palavras: “Bem, acho que ela não acreditava que seu raciocínio deveria ser consistente sob reflexão”, mas continuo tendo dificuldade em lidar com isso.

Consigo perceber um padrão nas palavras que saem de seus lábios, mas não consigo compreender a mente por trás delas em um nível empático. Consigo me imaginar no lugar de [alienígenas comedores de bebês](#) e da [Lady 3ª Kiritsugu](#), mas não consigo imaginar como é ser ela. Ou talvez eu simplesmente não queira?

É por isso que pessoas inteligentes têm apenas um certo tempo (medido em termos subjetivos de reflexão sobre religião) para se tornarem ateus. Após um determinado ponto, se você for perspicaz, tiver dedicado tempo pensando e defendendo sua religião, e ainda não se libertou das garras da [Epistemologia do Lado Negro](#), o interior da sua mente se torna como uma pintura de Escher.

(Um dos outros raros momentos em que ela parou — mencionei isso, caso você tenha oportunidade de usá-lo — foi quando ela estava falando sobre como é bom acreditar que alguém se importa se você faz o certo ou o errado — não, é claro, falando sobre como realmente existe um Deus que se importa se você faz o certo ou o errado; essa proposição não faz parte de sua religião...

E eu disse: “Mas eu me importo se você faz o certo ou o errado. Então, o que você está dizendo é que isso não é suficiente, e você também precisa acreditar em algo além da humanidade, que se importa se você faz o certo ou o errado.” Isso a fez parar um pouco, porque é claro que ela nunca havia pensado nisso nesses termos antes. (Foi apenas uma [ferramenta padrão da caixa de ferramentas fora do padrão](#).)

Mais tarde, em determinado momento, perguntei a ela se seria bom fazer algo diferente se Deus definitivamente não existisse, e desta vez ela respondeu: “Não”. “Então”, eu disse incrédulo, “se Deus existe ou não, isso não tem absolutamente nenhum efeito sobre como seria bom para as pessoas pensarem ou agirem? Acho que até mesmo um rabino ficaria um pouco desconfiado com isso.”

Sua religião parece agora consistir inteiramente na adoração da adoração. Assim como os verdadeiros crentes dos tempos antigos acreditavam que um pai todo-poderoso os salvaria, ela agora acredita que a crença em Deus a salvará.

Depois que ela disse: “Acredito que as pessoas são mais legais do que são”, perguntei: “Então, você sempre fica surpresa quando as pessoas superam suas expectativas?” Houve um longo silêncio e então, lentamente: “Bem... fico surpresa quando as pessoas... superam minhas expectativas?”

Na época, eu não compreendi essa pausa. Minha intenção era sugerir que, se ela constantemente se decepcionava com a realidade, então essa era uma desvantagem de acreditar falsamente. No entanto, ela parecia, em vez disso, surpresa com as implicações de não ser surpreendida.

Agora percebo que a essência de sua filosofia reside em sua crença de que ela havia se enganado, e a possibilidade de suas estimativas sobre outras pessoas serem de fato precisas, ameaçava a [Epistemologia do Lado Negro](#) que ela havia construído em torno de crenças como “Eu me benefico de acreditar que as pessoas

---

35 NT.: Texto original em inglês. *If there were a verb meaning “to believe falsely,” it would not have any significant first person, present indicative.*

36 NT.: Texto original em inglês. *“I reckon Granny ought to have a good look at you, Walter,” said Nanny. “I reckon your mind’s all tangled up like a ball of string what’s been dropped.”*

são mais amáveis do que realmente são”.

Ela retirou o antigo ídolo de seu trono e o substituiu por uma adoração explícita à Epistemologia do Lado Negro que havia sido inventada para defender o ídolo; ela adora sua própria tentativa de autoengano. Embora a tentativa tenha falhado, ela sinceramente não sabe disso.

E assim, os guardiões simbólicos da sanidade da humanidade (cujo lema é “estragar a sua festa insana desde os tempos de Epicuro”) agora precisam lutar contra a adoração ativa do autoengano — a adoração dos supostos benefícios da fé, em vez de Deus.

Na verdade, isso explica um fato sobre mim que eu nunca havia compreendido completamente antes: a razão pela qual fico irritado quando as pessoas falam como se o autoengano fosse fácil, e por que escrevo [ensaios inteiros](#) argumentando que fazer uma escolha deliberada de acreditar que o céu é verde é mais difícil de se libertar do que as pessoas parecem pensar. É porque, embora você não possa simplesmente escolher acreditar que o céu é verde, se você não reconhecer esse fato, poderá realmente enganar-se acreditando que teve sucesso em enganar a si mesmo.

E, uma vez que você sinceramente espera obter os benefícios que acredita que vêm do autoengano, acaba recebendo o mesmo tipo de benefício placebo que realmente viria de um autoengano bem-sucedido. Assim, ao explicar o quão difícil é o autoengano, estou, na verdade, direcionando-me diretamente aos benefícios-placebo que as pessoas obtêm ao acreditar que se enganaram e mirando nessa nova forma de religião que adora exclusivamente a adoração a Deus.

Eu me pergunto se essa batalha gerará uma nova lista de razões pelas quais não a crença em si, mas a crença na crença, é algo bom. Por que as pessoas obtêm grandes benefícios ao adorar sua própria adoração? Será que teremos que lidar novamente com a crença na crença na crença e com a adoração da adoração da adoração? Ou os teístas inteligentes finalmente desistirão desse tipo de argumentação?

Eu gostaria de poder acreditar que ninguém poderia acreditar na crença na crença na crença, mas o argumento do Mundo Zumbi na filosofia tornou-se ainda mais complexo do que isso, e seus defensores ainda não o abandonaram.

## Referências

- [1] Ludwig Wittgenstein, *Philosophical Investigations*, trans. Gertrude E. M. Anscombe (Oxford: Blackwell, 1953).
- [2] Terry Pratchett, *Maskerade*, Discworld Series (ISIS, 1997).

## 85 - O Paradoxo de Moore



O [Paradoxo de Moore](#) descreve a afirmação: “Está chovendo lá fora, mas eu não acredito que esteja”. Créditos a um [usuário do website MetaFilter](#), por esta explicação.

Minha compreensão do Paradoxo de Moore se aprofundou após ler comentários no *Less Wrong*. O [leitor](#) em seu comentário sugere que:

“Muitas pessoas não distinguem níveis de indireção. Para elas, “eu acredito em X” e “X” são a mesma coisa, e as razões para acreditar em X também justificam a veracidade de X.”

Discordo dessa avaliação. Crianças pequenas compreendem o conceito de crença falsa, o que implica separar mapa e território. Contudo, a ideia central é válida:

Muitas pessoas podem não diferenciar conscientemente entre acreditar e endossar algo.

Afinal, “Eu acredito na democracia” significa, coloquialmente, apoiar a democracia, não crer em sua existência. A palavra “crença” é ambígua. Talvez estejamos lidando com uma palavra confusa que causa (ou reflete) pensamento confuso.

No [exemplo original](#), ao dizer “Eu acredito que as pessoas são mais legais do que são”, a pessoa apresentou razões para crer na bondade das pessoas — benefícios para a saúde, etc. Ao experimentar um sentimento positivo ao “acreditar que as pessoas são legais”, ela refletiu sobre esse sentimento e concluiu: “Eu acredito que as pessoas são legais”. Ou seja, confundiu o sentimento positivo associado à crença com a crença em si. Simultaneamente, o mundo demonstrava que as pessoas não eram tão legais. Daí a afirmação paradoxal.

É fato que as pessoas são raramente ensinadas a reconhecer suas próprias crenças. Como na parábola do dragão invisível na garagem, quem afirma sua existência ignora a expectativa de não ver o dragão, indicando um modelo mental sem dragão.

Não há treinamento para reconhecer crenças. As escolas não ensinam: “Acreditar em algo significa que o mundo parece assim. Reconheça esse sentimento de crença real e diferencie-o de ter sentimentos positivos sobre uma crença que você reconhece como tal.”

[Isso elucidado](#) o caso real do Paradoxo de Moore e revela um mecanismo pelo qual as pessoas podem estar [simultaneamente certas e erradas](#).

Da mesma forma, [Kurige](#), que escreveu:

“Acredito que existe um Deus — que nos incutiu senso de certo e errado — e que a moralidade evoluiu em nós, resultado da formação de coalizões metapolíticas em comunidades de bonobos. Essas crenças não são contraditórias, mas a complexidade está em reconciliá-las.”

Suspeito, [Kurige](#), que você endossa a crença em Deus como fonte da moralidade e o veredicto da ciência. Ambas as comunidades parecem atraentes, com benefícios em seus conjuntos de crenças, e você se sente bem com ambas.

Mas você não disse:



“Deus nos inculcava senso de certo e errado, e a moralidade evoluiu em nós. Esses estados de realidade não são inconsistentes, mas a complexidade está em reconciliá-los.”

Se estiver lendo isso, Kurige, diga em voz alta a frase acima e perceba como ela soa menos aceitável — note a diferença subjetiva — antes de racionalizar novamente.

Essa é a diferença subjetiva entre ter motivos para apoiar duas crenças e seu modelo mental de um único mundo, uma única forma das coisas serem.

## 86 - Não acredite que você vai se autoenganar



Sem intenção de criticar Kurige, mas no *Less Wrong*, espere questionamentos se [disser](#):

Uma coisa que notei, e que ajuda a explicar a disparidade que sinto ao conversar com a maioria dos outros cristãos, é que, em algum momento, minha visão de mundo passou por uma grande mudança, afastando-se da fé cega e aproximando-se de um duplipensar orwelliano.

“Se você sabe que é um [duplipensar](#)... Como ainda pode acreditar nisso?” Eu me pergunto impotente. [Ou](#) então:

Escolhi acreditar na existência de Deus — de forma deliberada e consciente. No entanto, essa escolha não tem absolutamente nenhum impacto sobre a real existência de Deus.

Se você sabe que sua crença não está correlacionada com a realidade, como ainda pode acreditar? A percepção visceral de “Oh, espera, o céu realmente não é verde” não deveria ser seguida pela constatação de que “Meu mapa que afirma ‘o céu é verde’ não tem motivo para estar correlacionado com o território”?

Bem... aparentemente não.

Uma parte desse quebra-cabeça pode ser a minha explicação do [Paradoxo de Moore](#) (“Está chovendo lá fora, mas não acredito que esteja”) — as pessoas tendem a confundir, em sua introspecção, o afeto positivo associado a uma crença declarada com uma verdadeira crença.

Mas outra parte pode ser simplesmente que, ao contrário da indignação que inicialmente desejava expressar, é realmente muito fácil não fazer a transição de “O mapa que representa o território diria ‘X’” para realmente acreditar em “X”. Requer algum esforço explicar as ideias de como as mentes constroem a correspondência entre o mapa e o território e, mesmo assim, pode ser ainda mais difícil compreender essas implicações em um nível visceral.

Agora percebo que, quando escrevi “Você não pode acreditar que o céu é verde por um ato de vontade”, não estava apenas relatando fatos existentes imparcialmente. Eu também estava tentando estabelecer uma profecia autorrealizável.

Pode ser sensato sair repetindo deliberadamente: “Não consigo me safar com o pensamento duplo! No fundo, saberei que não é verdade! Se sei que meu mapa não está correlacionado com o território, significa que não acredito!”

Porque, dessa forma, se você for tentado a experimentar, os pensamentos de “Mas eu sei que isso não é verdade!” e “Não consigo me enganar!” surgirão prontamente em sua mente. Assim, você terá menos chances de se enganar com sucesso. É mais provável que você compreenda, instintivamente, que dizer a si mesmo “X” não torna “X” verdadeiro. Portanto, é realmente verdade que não-X.

Se você continuar repetindo para si mesmo que não pode simplesmente escolher acreditar deliberadamente que o céu é verde, será menos provável que caia em algum tipo de autoengano, seja acreditando verdadeiramente ou caindo no Paradoxo de Moore, na crença na crença ou na [crença no autoengano](#).

Se você continuar dizendo a si mesmo que no fundo você saberá...

Se você continuar afirmando, que apenas olhará para o seu mapa falsamente elaborado e saberá que

é um mapa sem qualquer correlação esperada com o território e, portanto, apesar de toda a elaboração, não investirá nenhuma credulidade nele...

Se você continuar dizendo a si mesmo, que a consistência reflexiva assumirá o controle e fará com que você pare de acreditar no nível do objeto, assim que chegar à percepção do meta nível de que o mapa não reflete a realidade...

Então, quando a tentação surgir, você pode, de fato, falhar.

Quando se trata de autoengano deliberado, acredite em sua própria incapacidade!

Afirme a si mesmo que o esforço está fadado ao fracasso — e assim será!

Esse é o poder do pensamento positivo ou do pensamento negativo?

De qualquer forma, isso me parece uma precaução sábia.



**I - Ver com novos olhos**



## 87 - Ancoragem e Ajuste



Imagine que eu gire uma Roda da Fortuna enquanto você observa, e ela pare no número 65. Então, daí eu te pergunto: a porcentagem de países africanos na ONU está acima ou abaixo desse número? Qual você acha que é essa porcentagem? Pense um pouco sobre essas perguntas, se quiser, mas sem pesquisar no Google.

Agora, tente adivinhar, em cinco segundos, o resultado da seguinte multiplicação.

Cinco segundos! Preparado? Vai!

$$1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$$

Tversky e Kahneman registraram as estimativas de pessoas que viram a Roda da Fortuna parar em diferentes números [1]. A média das estimativas de quem viu o número 65 foi 45%, enquanto a média de quem viu o número 10 foi 25%.

A teoria atual para esse tipo de experimento é que os participantes usam o número inicial, mesmo irrelevante, como ponto de partida, ou âncora. Depois, ajustam para cima ou para baixo até chegarem a uma resposta plausível, e param por aí. Geralmente, o ajuste é insuficiente, pois números mais distantes também podem ser plausíveis, mas paramos na primeira resposta satisfatória. Por exemplo, estudantes que viram “ $1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$ ” estimaram, em média, 512, enquanto os que viram “ $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$ ” estimaram 2.250. A hipótese é que os alunos tentaram multiplicar (ou adivinhar) os primeiros fatores e depois ajustaram para cima. Em ambos os casos, os ajustes foram insuficientes em relação ao valor real de 40.320, mas o primeiro grupo errou mais porque partiu de uma âncora menor.

Tversky e Kahneman relataram que oferecer recompensas pela precisão não diminui o efeito da ancoragem.

Strack e Mussweiler perguntaram em que ano Einstein visitou os Estados Unidos pela primeira vez [2]. Âncoras totalmente absurdas, como 1215 ou 1992, tiveram efeitos tão grandes quanto âncoras mais plausíveis, como 1905 ou 1939. Isso tem implicações claras em negociações salariais ou na compra de um carro.

Não sugiro que você explore esse fenômeno, mas esteja atento a quem o faz.

Observe seu próprio pensamento e tente perceber quando está ajustando um valor para chegar a uma estimativa.

Manipulações para evitar a ancoragem geralmente não são muito eficazes. Sugiro duas abordagens: primeiro, se o palpite inicial parecer absurdo, tente descartá-lo e fazer uma nova estimativa, em vez de ajustar a partir da âncora original. No entanto, isso pode não ser suficiente, pois participantes instruídos a evitar a ancoragem ainda a utilizam [3]. Portanto, em segundo lugar, mesmo tentando o primeiro método, pense também em uma âncora na direção oposta — uma âncora claramente muito pequena ou muito grande, em vez de apenas um pouco maior ou menor — e reflita brevemente sobre ela.

## Referências

- [1] Amos Tversky and Daniel Kahneman, "Judgment Under Uncertainty: Heuristics and Biases," *Science* 185, no. 4157 (1974): 1124–1131, doi:10.1126/science.185.4157.1124.
- [2] Fritz Strack and Thomas Mussweiler, "Explaining the Enigmatic Anchoring Effect: Mechanisms of Selective Accessibility," *Journal of Personality and Social Psychology* 73, no. 3 (1997): 437–446.
- [3] George A. Quattrone et al., "Explorations in Anchoring: The Effects of Prior Range, Anchor Extremity, and Suggestive Hints" (Unpublished manuscript, Stanford University, 1981).

## 88 - Pré-ativação e Contaminação



Imagine um experimento em que você deve apertar um botão se uma sequência de letras formar uma palavra e outro se não formar (por exemplo, “banack” versus “banner”). Em seguida, você vê a sequência “água”. Mais tarde, você identificará mais rapidamente “beber” como uma palavra. Isso se chama “pré-ativação cognitiva” ou *priming*; essa forma específica é a “pré-ativação semântica” ou “pré-ativação conceitual”.

O fascinante da pré-ativação é que ela ocorre ao nível básico — ela acelera o reconhecimento de letras como formadoras de uma palavra, algo que se espera que aconteça antes mesmo de pensarmos no significado da palavra.

A pré-ativação também revela o amplo paralelismo da propagação da ativação: ao ver “água”, provavelmente ativamos “rio”, “copo” ou “respingo”... e essa ativação se espalha, da conexão semântica de conceitos ao reconhecimento de sequências de letras.

A pré-ativação é subconsciente e inevitável, resultado da arquitetura neural humana. Tentar impedi-la conscientemente seria como tentar interromper a propagação da ativação em nossos próprios circuitos neurais. Experimente dizer em voz alta a cor — não o significado, mas a cor — da seguinte sequência de letras:

**VERDE**

No experimento de Mussweiler e Strack, os participantes receberam uma [pergunta-âncora](#): “A temperatura média anual na Alemanha é maior ou menor que 5 °C / 20 °C? [1]” Quem recebeu a âncora baixa foi mais rápido em identificar palavras como “frio” e “neve”, enquanto quem recebeu a âncora alta foi mais rápido em identificar “quente” e “sol”. Isso demonstra um mecanismo de não ajuste para ancoragem: a preparação de pensamentos e memórias compatíveis.

O resultado mais geral é que “informações” pseudo informativas, sabidamente falsas ou irrelevantes podem influenciar estimativas e decisões. No campo das heurísticas e vieses, esse fenômeno mais amplo é conhecido como contaminação [2].

Pesquisas iniciais sobre heurísticas e vieses revelaram os [efeitos da ancoragem](#), nos quais os participantes forneciam estimativas mais baixas (ou mais altas) da porcentagem de países africanos nas Nações Unidas, dependendo se inicialmente eram questionados se a porcentagem era maior ou menor que 10 (65). Originalmente, esse efeito foi atribuído aos participantes que se ajustavam a partir da âncora como ponto de partida, parando quando chegavam a um valor plausível e sub ajustando por pararem em uma extremidade do intervalo de confiança [3].

A hipótese inicial de Tversky e Kahneman ainda parece ser uma explicação correta em certas circunstâncias, principalmente quando os próprios participantes geram a estimativa inicial [4]. No entanto, pesquisas modernas indicam que a maioria da ancoragem é de fato devida à contaminação, e não ao ajuste deslizando. (Agradeço ao [usuário](#) que me lembrou disso — li o artigo de Epley e Gilovich há anos, em um capítulo sobre Heurísticas e Vieses, mas havia esquecido.)

O supermercado da sua cidade provavelmente exibe anúncios irritantes com dizeres como “Somente 12 por cliente” ou “Só 5 por US\$10”. Esses anúncios influenciam os clientes a comprar mais do que planejavam? Você provavelmente acredita que [não é afetado por eles](#). No entanto, alguns clientes devem ser afetados por eles, já que essas propagandas funcionam e é por isso que as lojas continuam fazendo uso deles [5].

No entanto, o aspecto mais preocupante da contaminação é que ela se torna mais uma das muitas faces do [viés de confirmação](#). Uma vez que uma ideia se instala em nossa mente, ela pré-ativa informações compatíveis para garantir sua existência contínua. Não importa quão fortes sejam as pressões para apresentar argumentos políticos, o viés de confirmação está enraizado em nosso *hardware* neural, com redes associativas preparando pensamentos e memórias alinhados. É um infeliz efeito colateral de nossa existência como criaturas neurais.

Uma única imagem fugaz pode ser suficiente para pré-ativar palavras associadas para o reconhecimento. Não pense que é necessário mais do que isso para ativar o viés de confirmação. Basta um breve instante e o [resultado final](#) já está decidido, pois [mudamos de opinião com menos frequência do que imaginamos...](#)

## Referências

- [1] Thomas Mussweiler and Fritz Strack, “Comparing Is Believing: A Selective Accessibility Model of Judgmental Anchoring,” *European Review of Social Psychology* 10 (1 1999): 135–167, doi:10.1080/14792779943000044.
- [2] Gretchen B. Chapman and Eric J. Johnson, “Incorporating the Irrelevant: Anchors in Judgments of Belief and Value,” in Gilovich, Griffin, and Kahneman, *Heuristics and Biases*, 120–138.
- [3] Tversky and Kahneman, “Judgment Under Uncertainty.”
- [4] Nicholas Epley and Thomas Gilovich, “Putting Adjustment Back in the Anchoring and Adjustment Heuristic: Differential Processing of Self-Generated and Experimentor-Provided Anchors,” *Psychological Science* 12 (5 2001): 391–396, doi:10.1111/1467-9280.00372.
- [5] Brian Wansink, Robert J. Kent, and Stephen J. Hoch, “An Anchoring and Adjustment Model of Purchase Quantity Decisions,” *Journal of Marketing Research* 35, no. 1 (1998): 71–81, <http://www.jstor.org/stable/3151931>.



## 89 - Acreditamos em tudo o que nos dizem?



Alguns experimentos iniciais sobre [ancoragem e ajuste](#) buscaram verificar se distrair os participantes — tornando-os cognitivamente “ocupados” ao pedir que se concentrassem no número “5” em sequências numéricas, por exemplo — reduziria o ajuste e, portanto, aumentaria a influência das âncoras. A maioria desses experimentos parecia confirmar a ideia de que a ocupação cognitiva aumentava a ancoragem e, de forma mais ampla, a [contaminação](#).

Ao examinar os resultados experimentais acumulados — com cada vez mais descobertas de contaminação, agravadas pela ocupação cognitiva — Daniel Gilbert identificou um padrão verdadeiramente surpreendente emergindo: será que acreditamos em tudo o que nos dizem?

Naturalmente, alguém poderia pensar que, ao ouvir uma afirmação, primeiro a compreenderíamos, depois a consideraríamos e, por fim, a aceitaríamos ou rejeitaríamos. Esse modelo aparentemente óbvio do fluxo do processo cognitivo remonta a Descartes. Porém, Spinoza, rival de Descartes, discordava dessa visão. Spinoza sugeriu que primeiro aceitamos passivamente uma afirmação enquanto a compreendemos e, somente depois, ativamente a rejeitamos por consideração.

Nos últimos séculos, os filósofos têm seguido majoritariamente a visão de Descartes, uma vez que sua perspectiva parecia mais lógica e intuitiva. No entanto, Gilbert viu uma maneira de testar experimentalmente as hipóteses de Descartes e Spinoza.

Se Descartes estiver correto, os participantes distraídos devem interferir tanto na aceitação de declarações verdadeiras quanto na rejeição de declarações falsas. Se Spinoza estiver correto, então os participantes distraídos devem lembrar-se de declarações falsas como verdadeiras, mas não devem lembrar-se de declarações verdadeiras como sendo falsas.

Gilbert, Krull e Malone confirmaram esse resultado, demonstrando que, entre os participantes para os quais foram apresentadas as novas declarações rotuladas como VERDADEIRAS ou FALSAS, a distração não teve efeito na identificação de declarações verdadeiras (55% de sucesso nas apresentações contínuas contra 58% nas interrompidas); porém, afetou a identificação de declarações falsas (55% de sucesso nas contínuas contra 35% nas interrompidas) [1].

Uma ilustração ainda mais dramática foi obtida em experimentos subsequentes realizados por Gilbert, Tafarodi e Malone [2]. Os participantes liam em voz alta relatórios de crimes que apareciam em um monitor de vídeo, sendo que a cor do texto indicava se uma determinada declaração era verdadeira ou **falsa**. Alguns relatórios continham declarações **falsas** que agravavam a gravidade do crime, enquanto outros continham declarações **falsas** que atenuavam ou desculpavam o crime. Além disso, alguns participantes também precisavam prestar atenção a sequências de números, procurando pelo número “5”, enquanto liam os relatórios criminais — essa tarefa de distração foi usada para criar ocupação cognitiva. Por fim, os participantes tiveram que recomendar a duração das penas de prisão para cada criminoso, variando de 0 a 20 anos.

Os indivíduos na condição de ocupação cognitiva recomendaram uma média de 11,15 anos de prisão para os criminosos da condição “agravante”, ou seja, criminosos cujos relatórios continham **declarações falsas rotuladas que agravavam a severidade do crime**. Já para os criminosos cujos relatórios continham **declarações falsas rotuladas que atenuavam o crime**, os participantes ocupados recomendaram uma média de 5,83 anos de prisão. Essa diferença quase dobrada foi estatisticamente significativa, como se poderia suspeitar.

Os participantes não ocupados leram os mesmos relatórios, com os mesmos rótulos e as mesmas sequências de números aparecendo ocasionalmente, exceto que eles não precisavam procurar pelo número “5”. Portanto, eles puderam dedicar mais atenção às declarações falsas rotuladas como “incrédulas”. Esses participantes não ocupados recomendaram uma média de 7,03 anos de prisão contra 6,03 anos para os criminosos cujos relatórios exacerbavam ou justificavam falsamente.

O artigo de Gilbert, Tatarodi e Malone tinha o título “Você não consegue não acreditar em tudo o que lê”.

Isso sugere, no mínimo, que devemos ter mais cuidado ao nos expor a informações não confiáveis, especialmente quando estamos fazendo outra coisa ao mesmo tempo. Portanto, é importante ter cuidado ao ler aquele jornal no supermercado.

PS: De acordo com um boato não verificado que acabei de inventar, as pessoas serão menos céticas em relação a este ensaio devido às mudanças de cor que distraem.

## Referências

[1] Daniel T. Gilbert, Douglas S. Krull, and Patrick S. Malone, “Unbelieving the Unbelievable: Some Problems in the Rejection of False Information,” *Journal of Personality and Social Psychology* 59 (4 1990): 601–613, doi:10.1037/0022-3514.59.4.601.

[2] Gilbert, Tatarodi, and Malone, “You Can’t Not Believe Everything You Read.”

## 90 - Pensamentos armazenados em cache



Um dos maiores mistérios sobre o cérebro humano é como ele consegue funcionar, sendo que a maioria dos neurônios dispara somente de 10 a 20 vezes por segundo, ou no máximo 200 Hz. Na neurologia, a “regra dos cem passos” estabelece que qualquer operação postulada deve ser concluída em no máximo 100 passos sequenciais. Pode-se ser tão paralelo quanto quiser, mas não se pode postular mais de 100 (preferencialmente menos) picos neurais em sequência.

Imagine ter que programar usando CPUs de 100 Hz, não importa quantas você tivesse. Seriam necessários cem bilhões de processadores apenas para fazer qualquer coisa em tempo real.

Se fosse preciso escrever programas em tempo real para cem bilhões de processadores de 100 Hz, um truque muito utilizado seria o armazenamento em cache. Isso envolve armazenar os resultados das operações anteriores e consultá-los na próxima vez, em vez de recalcular tudo do zero. É um processo muito neural, que envolve reconhecimento, associação e completar padrões.

É um bom palpite que a maioria da cognição humana consista em pesquisas em cache.

Esse pensamento costuma passar pela minha cabeça em determinados momentos.

Lembro de uma história muito ilustrativa que eu achei que tinha guardado, mas não consegui encontrá-la novamente: um homem tinha um vizinho que, certa vez, afirmou que a melhor maneira de remover uma chaminé era derrubar a lareira, esperar que os tijolos descessem um nível, derrubá-los e repetir o processo até que a chaminé desaparecesse. Anos depois, quando o homem precisou remover sua própria chaminé, esse pensamento oculto estava lá, à espreita, esperando para atacar...

Como percebeu o homem depois — você pode imaginar que não deu muito certo — seu vizinho não era particularmente conhecedor desses assuntos, não era uma fonte confiável. Se tivesse questionado a ideia, teria provavelmente percebido que ela era ruim. Às vezes, é melhor limpar o cache e recomeçar. O cérebro, todavia, completa o padrão automaticamente — e se você não perceber conscientemente que o padrão precisa de correção, você ficará com um padrão já completo.

Suspeito que se a ideia tivesse ocorrido ao homem — se ele próprio tivesse tido a brilhante ideia de como remover uma chaminé — ele teria examinado a ideia de maneira mais crítica. Mas quando outra pessoa já pensou em algo, podemos economizar poder de computação armazenando em cache a conclusão de outra pessoa, não é?

Na civilização moderna, em particular, ninguém pode pensar rápido o suficiente para repensar os seus próprios pensamentos. Se eu tivesse sido abandonado na floresta quando criança, criado por lobos ou robôs silenciosos, eu seria dificilmente reconhecido como um ser humano. Ninguém pode pensar rápido o suficiente para recapitular a sabedoria adquirida por uma tribo de caçadores-coletores numa vida inteira, começando do zero. Quanto à sabedoria de uma civilização alfabetizada, esqueça.

No entanto, noto com frequência que pessoas que buscam o pensamento crítico acabam repetindo ideias prontas que não foram criadas por pensadores críticos.

Um exemplo interessante é quando um cético afirma: “Bem, você não pode provar ou refutar uma religião por evidências factuais”. Já mencionei em outro lugar que isso é simplesmente falso se considerarmos a teoria da probabilidade. Além disso, isso também é falso em relação à verdadeira psicologia da religião — há

alguns séculos, dizer isso poderia ter te levado à fogueira. Por exemplo, uma mãe cuja filha tem câncer pode orar: “Deus, por favor, cure minha filha”, e não: “Querido Deus, eu sei que as religiões não podem ter consequências falsificáveis, o que significa que você não pode curar minha filha, então... bem, basicamente, estou rezando para me sentir melhor, em vez de fazer algo que possa realmente ajudar minha filha.”

Frequentemente, as pessoas leem a afirmação “Você não pode provar ou refutar uma religião por evidências factuais” e, então, quando se deparam com evidências que refutam uma religião, seu cérebro completa o padrão. Mesmo alguns ateus repetem esse absurdo sem hesitar. Se tivessem pensado na ideia por si próprios, em vez de apenas a ouvir de outra pessoa, teriam sido mais céticos.

Morte. Complete o padrão: “A morte dá sentido à... vida.” %

É frustrante quando conversamos com pessoas boas e decentes, que nunca pensariam em acabar com a espécie humana, e ao mencionarmos o risco existencial, ouvimos delas: “Bem, talvez a espécie humana não mereça sobreviver”. Elas nunca atirariam em seu próprio filho, que também é humano, mas o cérebro acaba completando esse padrão.

Quais padrões estão se completando na sua mente, sem você ter escolhido conscientemente?

Racionalidade. Complete o padrão: “O amor não é... racional.”

Se essa ideia surgisse de repente, como um pensamento completamente novo para você, como você a analisaria de forma crítica? Eu sei o que eu diria, mas o que você diria? É difícil olhar para as coisas de uma nova perspectiva. Tente evitar que sua mente complete o padrão da maneira usual, sem o surpreender. Talvez não haja uma resposta melhor do que a tradicional, mas você não conseguirá pensar nela se não impedir que sua mente a preencha automaticamente.

Após ler isso, quando você ouvir alguém repetindo sem hesitação um meme que considera falso ou sem sentido, talvez você se lembre de “Pensamentos armazenados em cache”. Minha crença agora está presente na sua mente, pronta para completar o padrão. Mas é verdade? Não deixe sua mente completar o padrão! Pense!

## 91 - A Caixa “Fora da Caixa”



Sempre que alguém o exorta a “pensar fora da caixa”, geralmente eles apontam, convenientemente, onde exatamente está esse “fora da caixa”. Não é engraçado como os inconformistas se vestem todos de maneira igual?...

Na área de Inteligência Artificial, todos que não são da área têm uma [ideia pré-concebida](#) de uma nova e revolucionária abordagem de IA: redes neurais, que funcionam exatamente como o cérebro humano! Uma nova ideia em IA. Complete o padrão: “As IAs lógicas, apesar de todas as grandes promessas, falharam em fornecer inteligência real por décadas — o que precisamos são redes neurais!”

Esse pensamento pré-concebido existe há três décadas. Ainda não alcançamos a inteligência geral. No entanto, de alguma forma, todos fora do campo sabem que as redes neurais são a nova ideia dominante que derrubou o paradigma desde que a retropropagação foi inventada na década de 1970. Fala sério, são como hippies envelhecidos.

Por sua própria natureza, as imagens inconformistas não permitem desvios da norma. Se você não veste preto, como as pessoas vão saber que você é um artista torturado? Como as pessoas reconhecerão sua singularidade se você não se encaixar no padrão do que é ser singular? Como alguém vai perceber que você tem um conceito revolucionário de IA, se não envolver redes neurais?

Outro exemplo desse mesmo equívoco é a “literatura subversiva”, que parece mais do mesmo, apoiada por um pequeno grupo desafiador de rebeldes que controlam todo o Departamento de Inglês. Como o usuário *Anonymous* [pergunta](#) no blog de Scott Aaronson:

“Alguma das ‘literaturas subversivas’ que você leu já fez com que você modificasse alguma de suas visões políticas?”

Ou como o usuário *Lizard* [observa](#):

“A revolução já foi televisionada. A revolução foi comercializada. A revolução se tornou uma mercadoria, um estilo de vida embalado, disponível no shopping local. Por US\$19,95, você obtém a máscara preta, a lata de spray, o cartaz ‘Esmague os fascistas’ e acesso ao seu blog, onde você pode escrever sobre a brutalidade policial que sofreu quando estava acorrentado a um hidrante. O capitalismo aprendeu a vender o anticapitalismo.”

Muitos no Vale do Silício observaram que a grande maioria dos investidores de risco, em determinado momento, estão todos perseguindo a mesma Inovação Revolucionária, aquela que teve um IPO seis meses atrás. Essa é uma observação particularmente esmagadora no mundo dos investimentos de risco por haver um motivo econômico direto para não seguir o rebanho — ou porque alguém já está desenvolvendo o produto ou porque alguém está apostando demais na startup. Steve Jurvetson uma vez me disse que na *Draper Fisher Jurvetson*, apenas dois parceiros precisam concordar para financiar qualquer startup em até US\$1,5 milhão. E se todos os parceiros concordarem que algo parece uma boa ideia, eles não farão isso. Se ao menos os comitês de concessão fossem tão sensatos.

O problema com a originalidade é que você realmente precisa pensar para alcançá-la, em vez de deixar seu cérebro preencher um padrão. Não há um “fora da caixa” convenientemente rotulado para onde você possa correr imediatamente. Há uma qualidade quase zen nisso — é como se você não pudesse ensinar o satori em palavras, porque o satori é a experiência das palavras falhando em transmitir seu significado.

Quanto mais você tentar seguir as instruções do Mestre Zen em palavras, mais distante estará de alcançar uma mente vazia.

Existe uma razão, em minha opinião, pela qual as pessoas não conseguem alcançar a novidade lutando por ela. Propriedades como verdade ou bom design são independentes da novidade:  $2 + 2 = 4$ , sim, isso é verdade, embora seja o que todo mundo também pensa. Aqueles que se esforçam para descobrir a verdade ou criar bons designs podem, com o tempo, alcançar a criatividade. Nem toda mudança é uma melhoria, mas toda melhoria é uma mudança.

Toda melhoria é uma mudança, mas nem toda mudança é uma melhoria. Aqueles que dizem “Quero criar uma ratoeira original!” e não “Quero criar uma ratoeira ideal!” geralmente desejam ser percebidos como originais. A “originalidade” nesse sentido é inerentemente social, porque só pode ser determinada por meio da comparação com outras pessoas. Assim, seu cérebro simplesmente preenche o padrão do que é percebido como “original”, e seus amigos concordam com a cabeça e dizem ser subversivo.

Os livros de negócios sempre informam, para sua conveniência, para onde seu queijo foi movido. Caso contrário, os leitores ficariam por aí se perguntando: “Onde está esse ‘Fora da Caixa’ que devo encontrar?”

Na verdade, o pensamento, assim como o satori, é um ato mental sem palavras.

Os eminentes filósofos do *Monty Python* expressaram isso brilhantemente em *Life of Brian* (A Vida de Brian) [1]:

“Vocês têm que pensar por vocês mesmos!”

“ Vocês são todos indivíduos!”

“Sim, somos todos indivíduos!”

“Vocês são todos diferentes!”

“Sim, somos todos diferentes!”

“Vocês todos têm que resolver isso vocês mesmos!”

“Sim, temos que resolver isso por nós mesmos!”

## Referências

[1] Graham Chapman et al., *Monty Python’s The Life of Brian* (of Nazareth) (Eyre Methuen, 1979).

## 92 - Visão original



Como Robert Pirsig colocou de forma tão perspicaz — e vou apenas o citar aqui, não sei se a história é factual ou não, mas de qualquer forma, é verdadeira [\[1\]](#).

Ele estava tendo problemas com alunos que não tinham nada a dizer.

No início, Pirsig pensou que fosse preguiça, mas logo ficou claro que não era o caso. Eles simplesmente não conseguiam pensar em nada para dizer.

Uma aluna em particular, usando óculos de lentes grossas, queria escrever um ensaio de quinhentas palavras sobre os Estados Unidos. Pirsig, acostumado com a frustração que surge a partir de declarações como essa, sugeriu, sem menosprezo, que ela reduzisse o tema para apenas a cidade de Bozeman, Montana.

Quando o prazo para a entrega do trabalho venceu, ela não havia conseguido terminar e ficou extremamente chateada. Ela tentou e tentou, mas simplesmente não conseguia pensar em nada para dizer.

Isso o deixou ainda mais perplexo. Agora, ele próprio não conseguia pensar em nada para dizer. Houve um momento de silêncio e, em seguida, ele deu uma resposta peculiar: “Reduza o tema para apenas a rua principal de Bozeman.” Foi um momento de compreensão. Ela assentiu obedientemente e saiu.

Porém, pouco antes da aula seguinte, ela voltou verdadeiramente angustiada, desta vez em lágrimas, o que deixou claro que essa angústia já estava presente fazia muito tempo. Ela ainda não conseguia pensar em nada para dizer e não conseguia entender por que, se não conseguia pensar em nada sobre a cidade de Bozeman na sua totalidade, deveria conseguir pelo menos pensar em algo para escrever sobre apenas uma rua da cidade.

Ele ficou furioso. “Você não está olhando de verdade!” ele exclamou. Uma lembrança retornou de quando ele mesmo foi demitido da universidade por ter coisas demais a dizer. Para cada fato, há uma infinidade de hipóteses. Quanto mais você observa, mais você vê. Ela realmente não estava observando e, de alguma forma, não conseguia entender isso.

Com raiva, ele disse: “Reduza o tema para a fachada de um prédio na rua principal de Bozeman. A *Opera House*. Comece com o tijolo superior esquerdo.”

Seus olhos, por trás dos óculos de lentes grossas, se arregalaram.

Na aula seguinte, ela entrou com uma expressão perplexa e entregou a ele um ensaio de cinco mil palavras escrito em frente à *Opera House*, na rua principal de Bozeman, Montana. “Sentei-me no café do outro lado da rua”, disse ela, “e comecei a escrever sobre o primeiro tijolo, e o segundo tijolo, e então, no terceiro tijolo, tudo começou a fluir e eu não conseguia parar. As pessoas acharam que eu estava louca e ficaram tirando sarro de mim, mas aqui está tudo. Eu não entendo o que aconteceu.”

Ele também não entendia, mas durante longas caminhadas pelas ruas da cidade, ele refletiu sobre o assunto e concluiu que ela estava claramente bloqueada pelo mesmo tipo de bloque-

io que o havia paralisado em seu primeiro dia de ensino. Ela estava bloqueada porque tentava repetir, em sua escrita, coisas que já havia ouvido, assim como no primeiro dia ele tentou repetir coisas que já havia decidido dizer. Ela não conseguia pensar em nada para escrever sobre Bozeman porque não conseguia se lembrar de nada que tivesse ouvido que valesse a pena repetir. De forma estranhamente inconsciente, ela não percebia que podia olhar e ver por si mesma, como escreveu, sem se preocupar com o que havia sido dito anteriormente. A redução do tema para um único tijolo destruiu o bloqueio, porque era tão evidente que ela precisava ter uma visão original e direta.

— Robert M. Pirsig,

*Zen and the Art of Motorcycle Maintenance* (Zen e a Arte da Manutenção de Motocicletas)

## Referências

[1] Pirsig, Zen and the Art of Motorcycle Maintenance.



## 93 - A realidade é mais estranha que a história



Suponha que eu lhe dissesse saber com absoluta certeza que as seguintes declarações eram verdadeiras:

- Se você se pintar com uma cor específica, exatamente entre o azul e o verde, isso inverterá a força da gravidade sobre você e fará com que você caia para cima.
- No futuro, o céu será preenchido por bilhões de esferas negras flutuantes. Cada esfera será maior do que todos os zepelins que já existiram juntos. Se você oferecer dinheiro a uma esfera, ela baixará um prostituto do céu por meio de uma corda elástica.
- Seus netos acreditarão que não apenas é tolice, mas também maldade colocar ladrões na prisão em vez de espancá-los.

Você provavelmente me consideraria louco, não é mesmo?

Agora, imagine-se no ano de 1901, tendo que escolher entre acreditar nas afirmações que acabei de fazer e acreditar em declarações como as seguintes:

- Existe um limite absoluto para a velocidade com que dois objetos parecem se mover um em relação ao outro, o qual é exatamente 670.616.629,2 milhas por hora (1.079.251.985,1 quilômetros por hora). Se você estiver em um trem quase tão rápido e disparar um tiro com uma arma pela janela, as unidades fundamentais de comprimento mudam e a bala parecerá acelerar à sua frente, mas outras pessoas verão algo diferente. Ah, e o tempo também muda.
- No futuro, haverá uma rede global superconectada de bilhões de máquinas de cálculo, cada uma com mais poder do que todas as máquinas de cálculo anteriores a 1901 juntas. Uma das principais utilidades dessa rede será transmitir imagens em movimento de sexo lésbico, fingindo que essas imagens são feitas apenas de números.
- Seus netos acreditarão que não apenas é tolo, mas também cruel afirmar que alguém não deveria ser presidente dos Estados Unidos por ser negro.

Com base em um [comentário](#) de Robin Hanson: *“Fico imaginando se alguém poderia descrever com detalhes suficientes uma história fictícia de uma realidade alternativa, uma realidade que nossos ancestrais seriam incapazes de distinguir da verdade, a fim de destacar o quão surpreendente a verdade se tornou.”*

## 94 - A falácia lógica da generalização a partir de evidências fictícias



Quando tento apresentar o assunto de IA avançada, qual é a primeira coisa que ouço, mais da metade do tempo?

“Ah, você quer dizer como nos filmes do Exterminador do Futuro, Matrix e Robôs do Asimov!”

E eu respondo: “Bem, não, não exatamente. Tento evitar a falácia lógica de generalizar a partir de evidências fictícias.”

Algumas pessoas entendem imediatamente e riem. Outras defendem o uso do exemplo, discordando que seja uma falácia.

Qual é o problema em usar filmes ou romances como ponto de partida para a discussão? Afinal, ninguém está afirmando que isso é verdade. Onde está a falsidade, onde está o pecado do racionalismo? A ficção científica é a tentativa do autor de visualizar o futuro; por que não aproveitar o pensamento que já foi feito em nosso nome, em vez de começar do zero?

Nem todo deslize na dança precisa da racionalidade requer necessariamente uma crença absoluta em uma falsidade; existem maneiras mais sutis de cometer erros.

Primeiro, abandonemos a ideia de que a ficção científica representa uma tentativa racional de prever o futuro. Até os escritores de ficção científica mais diligentes são, acima de tudo, contadores de histórias; as exigências de contar histórias não são as mesmas que as exigências de previsão. Como [observa](#) Nick Bostrom [\[1\]](#):

Quando foi a última vez que você assistiu a um filme sobre a extinção repentina da humanidade (sem aviso e sem substituição por outra civilização)? Embora esse cenário possa ser muito mais provável do que um cenário em que heróis humanos repelem com sucesso uma invasão de monstros ou guerreiros robôs, não seria tão divertido de se assistir<sup>37</sup>.

Portanto, existem [distorções específicas](#) na ficção. No entanto, tentar corrigir essas distorções específicas, não é o suficiente. Uma história nunca é uma tentativa racional de análise, nem mesmo para os escritores de ficção científica mais diligentes, pois as histórias não utilizam distribuições de probabilidade. Vou ilustrar da seguinte maneira:

Bob Merkelthud deslizou cuidadosamente pela porta da nave espacial alienígena, olhando primeiro para a direita e depois para a esquerda (ou vice-versa), para ver se algum dos temidos Monstros Espaciais ainda estava presente. Ao seu lado, havia a única arma considerada eficaz contra os Monstros Espaciais: uma Espada Espacial forjada em titânio puro com uma probabilidade de 30%, um pé de cabra comum de ferro com uma probabilidade de 20% e um disco preto brilhante encontrado nas ruínas fumegantes de *Stonehenge* com uma probabilidade de 45%, sendo os 5% restantes distribuídos em vários resultados menores que não valem a pena listar aqui.

---

37 NT.: Texto original em inglês. *When was the last time you saw a movie about humankind suddenly going extinct (without warning and without being replaced by some other civilization)? While this scenario may be much more probable than a scenario in which human heroes successfully repel an invasion of monsters or robot warriors, it wouldn't be much fun to watch.*

Merkelthud (embora houvesse uma chance significativa de Susan Wiffefoofer estar lá em seu lugar) deu dois passos para frente ou um para trás, quando um grande rugido quebrou o silêncio da câmara escura! Ou talvez fosse o zumbido silencioso do fundo da câmara de descompressão branca! Embora Amfer e Woofi (1997) argumentem que Merkelthud seja devorado neste ponto, Spacklebackle (2003) aponta que—

Os personagens podem ser ignorantes, mas o autor não pode simplesmente dizer as três palavras mágicas “eu não sei”. O protagonista deve traçar uma linha única em direção ao futuro, repleta de detalhes que dão vida à história, desde as atitudes futuristas apropriadas de Wiffefoofer em relação ao feminismo até a cor de seus brincos.

Em seguida, todos esses detalhes complicados e suposições questionáveis são agrupados e recebem um rótulo curto, criando a ilusão de que são um [único conjunto](#).

Em problemas com grandes espaços de respostas, a maior dificuldade não é verificar a resposta correta, mas simplesmente [localizá-la no espaço de respostas em primeiro lugar](#). Se alguém começa perguntando se as IAs vão ou não nos colocar em cápsulas como em “Matrix”, estão pulando para uma proposição de 100 bits, sem os 98 bits correspondentes de evidência para localizá-la no espaço de respostas como uma possibilidade digna de consideração explícita. Seriam necessárias apenas algumas evidências adicionais após os primeiros 98 bits para promover essa possibilidade para quase certeza, o que revela onde a maior parte do trabalho é realizada.

A etapa “preliminar” de localizar possibilidades dignas de consideração explícita inclui etapas como: avaliar o que você sabe e o que não sabe, o que você pode e o que não pode prever, fazer um esforço consciente para evitar o [viés do absurdo](#) e [ampliar os intervalos de confiança](#), ponderar quais perguntas são as mais importantes, tentar ajustar para possíveis Cisnes Negros e considerar incógnitas anteriormente desconhecidas. Pular para “Matrix: Sim ou Não?” [ignora tudo isso](#).

Qualquer negociador profissional sabe que controlar os termos de um debate é praticamente controlar o resultado do debate. Se você começa pensando em “Matrix”, isso sugere exércitos de robôs marchando e derrotando humanos após uma longa batalha — não uma superinteligência estalando os dedos com nanotecnologia. Ele se concentra em uma batalha “nós contra eles”, direcionando a atenção para questões como “quem vai vencer?” e “quem deveria vencer?” e “Será que as IAs realmente serão assim?” Isso cria uma atmosfera geral de entretenimento, de “Qual é a sua visão incrível do futuro?”

Perdas na ressonância vazia estão: considerações sobre mais de um projeto mental possível que uma “Inteligência Artificial” poderia implementar; a dependência do futuro das condições iniciais; o poder de uma inteligência mais avançada que a humana e o argumento de sua [imprevisibilidade](#); pessoas levando todo o assunto a sério e tentando fazer algo a respeito.

Se algum manipulador insidioso de debates decidisse que seu resultado preferido seria mais bem servido, forçando os debatedores a começarem refutando o “Exterminador do Futuro”, eles teriam sucesso em distorcer a questão. Ao debater sobre o controle de armas, o porta-voz da NRA não deseja ser retratado como um “maníaco por armas”, e o oponente do desarmamento não deseja ser apresentado como um “defensor desarmamentista das vítimas”. Por que você permitiria a mesma distorção de contexto por parte dos roteiristas de Hollywood, mesmo que acidentalmente?

Os jornalistas não me dizem: “O futuro será como ‘2001’”. Mas eles perguntam: “O futuro será como ‘2001’ ou será como a IA?” Esse é um problema de enquadramento tão grande quanto perguntar: “Devemos cortar benefícios para veteranos deficientes ou aumentar impostos para os ricos?”

No ambiente ancestral, não havia imagens em movimento; o que você via com seus próprios olhos era verdade. Um vislumbre momentâneo de uma única palavra pode nos [influenciar](#) e tornar os pensamentos mais disponíveis, como foi demonstrado em estimativas de probabilidade. Quanto dano você acha que um filme de duas horas pode causar em seu julgamento? Já é difícil desfazer o dano causado por concentração deliberada — por que convidar o vampiro para dentro de sua casa? No xadrez ou no Go, cada jogada desperdiçada é uma perda; na racionalidade, qualquer influência não baseada em evidências é, em média, entrópica.

Os espectadores de cinema conseguem [desacreditar](#) o que veem? Pelo que posso dizer, poucos espectadores de filmes agem como se tivessem testemunhado diretamente o futuro da Terra. As pessoas que assistiram aos filmes do “Exterminador do Futuro” não se esconderam em abrigos antiaéreos em 29 de agosto de 1997. Todavia, aqueles que cometem a falácia, parecem agir como se tivessem presenciado os eventos do filme ocorrendo em algum outro planeta; não na Terra, mas em algum lugar semelhante à Terra.

Você diz: “Suponha que construímos uma IA muito inteligente”, e eles dizem: “Mas isso não resultou em uma guerra nuclear no ‘Exterminador do Futuro?’” Pelo que posso dizer, é um raciocínio idêntico, até o tom de voz, daquele que poderia dizer: “Mas isso não resultou em uma guerra nuclear em Alpha Centauri?” ou “Isso não resultou na queda da cidade-estado italiana de Piccolo no século XIV?” O filme não é considerado uma profecia, mas sim um caso histórico ilustrativo. A história se repetirá? Quem sabe?

Em uma recente discussão sobre a explosão de inteligência, alguém mencionou que Vinge não parecia acreditar que as interfaces cérebro-computador aumentariam significativamente a inteligência, citando *Marooned in Realtime* (Naufragado em tempo real) e Tunç Blumenthal como exemplos em que o personagem mais avançado não parecia tão poderoso.

Eu respondi indignado: “Mas Tunç perdeu a maior parte de seu hardware! Ele era aleijado!” E então percebi meu erro e pensei comigo mesmo: “O que diabos estou dizendo? A questão não precisa ser discutida apenas com base em como Vinge retratou seus personagens? Tunç Blumenthal não é ‘aleijado’, ele é irreal. Eu poderia dizer: ‘Vinge escolheu retratar Tunç como aleijado, por razões que podem ou não estar relacionadas com sua melhor previsão pessoal’, e isso atribuiria ao seu aspecto autoral um peso de evidência adequado. Mas não posso dizer: ‘Tunç era aleijado’. Não há um passado para Tunç Blumenthal.”

Cometi deliberadamente um erro em meu primeiro rascunho deste ensaio ao afirmar: “Outros defendem o uso do exemplo, discordando que seja uma falácia”. Mas “Matrix” não é um exemplo!

Uma falha semelhante é a falácia lógica de argumentar com base em evidências imaginárias: “Bem, se você fosse até o fim do arco-íris, encontraria um pote de ouro - o que apenas comprova meu ponto!” (Atualizar evidências previstas, mas não observadas, é a imagem matemática espelhada do viés retrospectivo.)

O cérebro possui muitos mecanismos para generalizar a partir da observação, não apenas a heurística da disponibilidade. Quando você vê três zebras, forma a categoria “zebra”, e essa categoria incorpora uma inferência perceptiva automática. Criaturas em forma de cavalo com listras brancas e pretas são classificadas como “zebras”, portanto, espera-se que sejam rápidas e boas para comer; espera-se que sejam semelhantes a outras zebras observadas.

Da mesma forma, quando as pessoas veem (imagens em movimento de) três Borgs, seus cérebros criam automaticamente a categoria “Borg” e inferem automaticamente que os humanos com interfaces cérebro-computador pertencem à classe “Borg”, e que eles todos serão semelhantes aos outros Borgs observados: frios, sem compaixão, vestindo couro preto e andando com passos mecânicos pesados. Os jornalistas não acreditam que o futuro incluirá os Borgs - eles não consideram “Jornada nas Estrelas” como uma profecia. Mas quando alguém fala sobre interfaces cérebro-computador, eles pensam: “O futuro terá os Borgs?” Não pensam: “Como a telepatia assistida por computador torna as pessoas menos legais?” Não pensam: “Eu nunca vi um Borg e nem mais ninguém os viu”. Não, estão formando um estereótipo baseado em evidências literalmente nulas.

Como George Orwell [disse](#) sobre os clichês [2]:

Acima de tudo, é fundamental permitir que o significado escolha a palavra, e não o contrário. Quando você pensa em algo abstrato, tende a recorrer às palavras desde o início, e a menos que faça um esforço consciente para evitá-lo, o dialeto existente virá correndo e fará o trabalho por você, confundindo ou até mesmo alterando seu significado<sup>38</sup>.

---

38 NT.: Texto original em inglês. *What is above all needed is to let the meaning choose the word, and not the other way around... When you think of something abstract you are more inclined to use words from the start, and unless you make a conscious effort to prevent it, the existing dialect will come rushing in and do the job for you, at the expense of blurring or even changing your meaning.*

No entanto, em minha opinião, o aspecto mais prejudicial de usar a imaginação de outros autores é que isso impede as pessoas de usarem a sua própria. Como [disse](#) Robert Pirsig [3]:

Ela estava bloqueada porque tentava repetir, em sua escrita, coisas que já ouvira, assim como no primeiro dia ele tentara repetir coisas que já decidira dizer. Ela não conseguia pensar em nada para escrever sobre Bozeman porque não conseguia se lembrar de nada que tivesse ouvido que valesse a pena repetir. Ela estava estranhamente inconsciente de que podia olhar e ver por si mesma, como escreveu, sem considerar o que havia sido dito antes<sup>39</sup>.

As ficções lembradas invadem rapidamente e constroem o seu pensamento por você; elas substituem a visão — a conveniência mais mortal de todas.

## Referências

[1] Nick Bostrom, “Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards,” *Journal of Evolution and Technology* 9 (2002), <http://www.jetpress.org/volume9/risks.html>.

[2] Orwell, “Politics and the English Language.”

[3] Pirsig, *Zen and the Art of Motorcycle Maintenance*.

---

39 NT.: Texto original em inglês. *She was blocked because she was trying to repeat, in her writing, things she had already heard, just as on the first day he had tried to repeat things he had already decided to say. She couldn't think of anything to write about Bozeman because she couldn't recall anything she had heard worth repeating. She was strangely unaware that she could look and see freshly for herself, as she wrote, without primary regard for what had been said before.*

## 95 - A virtude da especificidade



O que é verdade para uma maçã pode não ser verdade para outra maçã; assim, mais pode ser dito sobre uma única maçã do que sobre todas as maçãs do mundo.

— As Doze Virtudes da Racionalidade

Dentro de suas próprias profissões, as pessoas entendem a importância da especificidade. Um mecânico de automóveis reconhece a diferença entre um carburador e um radiador e não os consideraria apenas como “peças de carro”. Um caçador-coletor distingue entre um leão e uma pantera. Um zelador não limpa o chão com um limpador de vidros, mesmo que as garrafas sejam parecidas para quem não domina essa arte.

No entanto, fora de suas próprias profissões, as pessoas muitas vezes cometem o erro de tentar ampliar uma palavra ao máximo, abrangendo o maior território possível. Será mais glorioso, mais sábio, mais impressionante falar sobre todas as maçãs do mundo? Quão mais elevado seria explicar o pensamento humano em geral, sem se preocupar com questões menores, como as técnicas para resolver um cubo de Rubik. Na verdade, parece quase desnecessário considerar questões específicas; uma teoria geral não é uma conquista digna o suficiente por si só?

É hábito dos curiosos levantar uma pedra entre um milhão de pedras na praia e encontrar algo novo nela, algo interessante, algo diferente. Eles chamam essas pedras de “diamantes” e se perguntam o que pode haver de especial nelas — quais qualidades internas podem compartilhar, além do brilho que notaram inicialmente. Então, alguém chega e diz: “Por que não chamar também essa pequena pedra de diamante? E esta aqui, e esta?”. Eles estão entusiasmados e têm boas intenções. Afinal, parece antidemocrático, excludente, elitista e não holístico chamar algumas pedras de “diamantes” e outras não. Parece... sem imaginação... se me permite dizer. Pouco aberto, pouco abrangente, pouco comunitário.

Pode parecer poético dar a uma palavra muitos significados e, assim, espalhar nuances de conotação por toda parte. No entanto, mesmo os poetas, se forem bons poetas, devem aprender a ver o mundo com precisão. Não é suficiente comparar o amor a uma flor. O amor ciumento e não consumado não é o mesmo que o amor de um casal casado há décadas. Se você precisa de uma flor para simbolizar o amor ciumento, deve ir ao jardim, observar e fazer distinções sutis — encontrar uma flor com perfume inebriante, uma cor brilhante e espinhos. Mesmo que sua intenção seja sombrear significados e lançar conotações, é necessário manter um registro preciso de exatamente quais significados estão sendo sombreados e conotados.

É uma parte necessária da arte do racionalista — ou mesmo da arte do poeta! — focar estritamente em pedras incomuns que possuem alguma qualidade especial. E observe os detalhes que essas pedras — e apenas essas pedras! — compartilham entre si. Isso não é um pecado.

É totalmente apropriado que os biólogos evolucionistas modernos expliquem apenas os padrões encontrados nas criaturas vivas e não a “evolução” das estrelas ou a “evolução” da tecnologia. Infelizmente, algumas pessoas usam a mesma palavra “evolução” para abranger os padrões de replicação selecionados naturalmente na vida, a estrutura acidental das estrelas e a estrutura inteligentemente projetada da tecnologia. E, como todos sabemos, se as pessoas usam a mesma palavra, então deve ser tudo a mesma coisa. Automaticamente generalizamos qualquer coisa que pensamos saber sobre a evolução biológica para a tecnologia. Qualquer um que nos diga o oposto é apenas um pedante inútil. É impensável que nossa ignorância da teoria evolucionária moderna seja tão grande que não possamos distinguir a diferença entre um carburador

e um radiador. Não, a outra pessoa — você sabe, aquela que estudou matemática - é burra demais para ver as conexões.

E o que poderia ser mais virtuoso do que ver conexões? Certamente, os mais sábios de todos os seres humanos são os gurus da Nova Era que dizem: “Tudo está conectado a tudo mais”. Se você disser isso em voz alta, faça uma pausa, para que todos absorvam o choque absoluto dessa Sabedoria Profunda.

Existe um mapeamento trivial entre um grafo e seu complemento. Um grafo totalmente conectado, com uma aresta entre cada par de vértices, transmite a mesma quantidade de informação que um grafo sem arestas. As redes importantes são aquelas nas quais algumas coisas não estão conectadas a outras.

Quando pessoas não iluminadas tentam ser profundas, elas traçam comparações verbais intermináveis entre este tópico e aquele tópico, que são assim ou assado, até que sua rede esteja totalmente, conectada e também totalmente inútil. O remédio é o conhecimento específico e o estudo. Quando você compreende as coisas em detalhes, pode ver como elas são diferentes e começar a remover entusiasticamente as arestas da sua rede

Da mesma forma, as categorias importantes são aquelas que não abrangem tudo no universo. Boas hipóteses podem explicar apenas alguns resultados possíveis e não outros.

Era perfeitamente correto para Isaac Newton explicar apenas a gravidade, o modo como as coisas caem, como os planetas orbitam o Sol e como a Lua gera as marés, mas não abordar o papel do dinheiro na sociedade humana ou como o coração bombeia o sangue. Zombar da estreiteza é bastante semelhante ao pensamento dos antigos gregos, que acreditavam que sair e observar as coisas era um trabalho manual, e o trabalho manual era destinado aos escravos.

Como Platão afirmou em “A República”, Livro VII [\[1\]](#):

Se alguém jogasse a cabeça para trás e aprendesse algo olhando para os vários padrões no teto, aparentemente você pensaria que ele estava contemplando com sua razão, quando ele estava apenas olhando... Não posso deixar de acreditar que nenhum estudo faz a alma olhar para o alto, exceto aquele que se preocupa com o ser real e o invisível. Se ele fica boquiaberto e olha para cima, ou fecha a boca e olha para baixo, se são coisas dos sentidos que ele tenta aprender algo, eu declaro que ele nunca poderia aprender, pois nenhuma dessas coisas admite conhecimento: eu digo que sua alma está olhando para baixo, não para cima, mesmo que esteja flutuando de costas na terra ou no mar!<sup>40</sup>

Muitos hoje cometem um erro semelhante e pensam que conceitos estreitos são tão inferiores, pouco elevados e pouco filosóficos quanto, digamos, sair e observar as coisas — um esforço adequado apenas para a classe baixa. Mas os racionalistas — e também os poetas — precisam de palavras precisas para expressar pensamentos precisos; eles precisam de categorias que incluam apenas algumas coisas e excluam outras. Não há nada de errado em focar a mente, restringir as categorias, excluir possibilidades e aprimorar as proposições. Na verdade, não há! Se tornarmos as palavras excessivamente amplas, acabaremos com algo que é falso e nem mesmo faz uma boa poesia.

*E nem me fale sobre as pessoas que pensam que a Wikipédia é uma “inteligência artificial”, que a invenção do LSD foi uma “singularidade” ou que as corporações são “superinteligentes”!*

## Referências

[1] Plato, Great Dialogues of Plato, ed. Eric H. Warmington and Philip G. Rouse (Signet Classic, 1999).

---

40 NT.: Texto original em inglês. *If anyone should throw back his head and learn something by staring at the varied patterns on a ceiling, apparently you would think that he was contemplating with his reason, when he was only staring with his eyes... I cannot but believe that no study makes the soul look on high except that which is concerned with real being and the unseen. Whether he gape and stare upwards, or shut his mouth and stare downwards, if it be things of the senses that he tries to learn something about, I declare he never could learn, for none of these things admit of knowledge: I say his soul is looking down, not up, even if he is floating on his back on land or on sea!*

## 96 - Como parecer (e ser) profundo



Recentemente, participei de um grupo de discussão sobre a Morte. Esse encontro despertou emoções profundas. Entre todos os almoços no Vale do Silício, este foi o mais sincero. As pessoas compartilhavam experiências sobre a morte de familiares e amigos. Também expressavam reflexões sobre sua própria mortalidade. O mais [impressionante](#) era como todos se ouviam atentamente. Fiquei curioso: como recriar essas condições em outras ocasiões?

No grupo, eu era o único transhumanista. Tomei cuidado para não parecer desagradável com essa perspectiva. Como dizem, “fanático é quem não muda de ideia nem de assunto”. Eu sempre me esforço para, pelo menos, mudar de assunto. As pessoas abordaram o significado que a morte dá à vida. Alguns a viam como uma bênção disfarçada. Com cautela, expliquei: [os transumanistas têm uma visão positiva da vida, mas desaprovam a morte.](#)

Após a discussão, várias pessoas me disseram que eu era “profundo”. Bem, sou, mas isso me fez refletir: o que faz alguém parecer profundo? Em certo momento, uma mulher compartilhou:

“Pensar na morte me leva a tratar as pessoas com simpatia. Quem sabe se as verei de novo? Quando tenho algo positivo para dizer, agora o digo imediatamente.”

Respondi: “Que bela reflexão. Se um dia nos livrarmos da ameaça da morte, espero que você ainda continue agindo assim...”

Essa mulher foi uma das que me chamaram de profundo.

Em outro momento, um homem mencionou um benefício específico da morte. Não me lembro qual. Eu disse: “Pense na natureza humana. Se as pessoas levassem pancadas semanais na cabeça, logo achariam isso bom. Mas se você perguntasse a alguém que não sofre isso se gostaria de experimentar, a resposta seria não. Da mesma forma, um imortal provavelmente não escolheria morrer por esse suposto benefício.”

Depois, esse homem também me chamou de profundo.

A correlação não implica causalidade. Talvez eu estivesse apenas falando com voz grave e soasse sábio. Mas suspeito que pareci “profundo” por desafiar consistentemente [o padrão](#) de “sabedoria profunda” de uma forma que fazia sentido imediato.

Existe um estereótipo de “Sabedoria Profunda” sobre a morte. Complete o padrão: “A morte dá sentido à vida”. Todos conhecem essa resposta padrão do “Sábio Profundo”. É como um aplauso automático. Quando mencionada, as pessoas concordam por reflexo. Podem até dizer “Que sabedoria profunda!”, talvez para parecerem profundas também. Mas não ficarão surpresas; não ouvirão nada novo ou [fora da caixa](#). Isso é “crença na sabedoria” — o pensamento é rotulado como “profundamente sábio”, mas não traz insights reais.

Pessoas que tentam parecer sábias muitas vezes soam vazias. Parecem apenas [ecoar](#), não ser autênticas, porque estão tentando demais ao invés de [otimizar](#) verdadeiramente seu pensamento.

Quanta reflexão eu precisei para parecer profundo?

O cérebro humano opera a apenas 100 Hz, e eu respondi em tempo real. A maior parte do trabalho deve ter sido pré-computada. O desafio foi escolher uma resposta compreensível em uma etapa e formulá-la



para máximo impacto.

Filosoficamente, muito do meu trabalho já estava feito. Complete o padrão: A condição X é justificada pelo benefício Y. “Seria uma falácia naturalista?” / “Seria um viés do status quo?” / “Seria possível obter Y sem X?” / “Se X fosse novidade, aceitaríamos para obter Y?” Executo esses padrões quase automaticamente, como respirar. Afinal, muito do pensamento humano deve ser automático para o cérebro funcionar.

Além disso, eu já defendia a filosofia transhumanista desenvolvida. Ela questiona os pensamentos sobre a morte. Complete o padrão: “A morte é uma tragédia sem sentido que as pessoas racionalizam”. Esse padrão era desconhecido pelos meus ouvintes. Tive várias chances de usá-lo, sempre ligado ao mesmo tema. Isso me deu coerência e originalidade.

Suspeito que seja por isso que a filosofia oriental parece profunda para os ocidentais. Ela oferece um refúgio coerente, mas fora do padrão, para a Sabedoria Profunda. De forma simétrica, em obras japonesas, às vezes [cristãos](#) são retratados como sábios ou místicos (outras vezes [não](#)).

Lembro-me de um economista que observou: o público conhece tão pouco de economia convencional que, na TV, ele só precisava repetir conceitos básicos para parecer brilhantemente original.

Foi crucial que meus ouvintes entendessem minha resposta imediatamente. Podiam não concordar, mas fazia sentido para eles. Conheço transumanistas incapazes de parecer profundos por não entenderem o que seu ouvinte ainda não sabe. Para parecer profundo, nunca diga nada que esteja acima ou além do estado mental atual do seu interlocutor. É assim que funciona.

Para parecer profundo:

Estude filosofias fora do comum.

Busque discussões que te permitam parecer profundo.

Pré-elabore seus pensamentos filosóficos para explicá-los claramente.

Acima de tudo, fique dentro do limite de uma única etapa inferencial.

Para ser verdadeiramente profundo:

Pense por si mesmo sobre assuntos “sábios”, importantes ou emocionais. Pensar por si mesmo não é o mesmo que pensar em [repostas incomuns](#). Veja por si mesmo, em vez de deixar seu cérebro completar o padrão.

Não se [contente com a primeira resposta](#).

Descarte ideias vagamente insatisfatórias.

Com o tempo, seus pensamentos formarão um todo coerente. Fluirão de uma única fonte, em vez de serem meras repetições das conclusões alheias.

## 97 - Mudamos de ideia com menos frequência do que pensamos



Nos últimos anos, fizemos um experimento discreto. Abordamos colegas que escolhiam entre diferentes ofertas de emprego. Pedimos que estimassem a probabilidade de escolher um emprego sobre outro. A confiança média na escolha prevista foi modesta: 66%. Mas apenas 1 dos 24 entrevistados escolheu a opção menos provável inicialmente. Isso resultou em uma precisão geral de 96%.<sup>41</sup>

— Dale Griffin e Amos Tversky [1]

Foi em 1º de agosto de 2003, por volta das 15h, que minha forma de pensar mudou. Percebi algo importante: uma vez que eu pudesse prever minha resposta, provavelmente já havia decidido. Mudamos de opinião menos do que imaginamos. Na maioria das vezes, adivinhamos nossa resposta meio segundo após ouvir a pergunta.

Quão rápido passa aquele momento despercebido! Aquele instante em que ainda não prevemos nossa resposta. É uma pequena janela de oportunidade para a inteligência agir. Isso vale tanto para escolhas quanto para fatos.

[O princípio do resultado final](#) é claro: apenas as verdadeiras causas de nossas crenças determinam sua eficácia como racionalistas. Uma vez estabelecida uma crença, nenhum argumento muda seu valor de verdade. Uma vez tomada uma decisão, nenhuma argumentação muda suas consequências.

Você pode acreditar que chega a uma crença por meios não racionais. Depois, tenta justificá-la. Se perceber que não consegue justificar, pode rejeitá-la.

No entanto, mudamos de ideia muito menos — muito menos mesmo — do que pensamos.

Tenho certeza de que você se lembra de ter mudado de ideia alguma vez. Todos nos lembramos. Mas e as vezes em que não mudamos? Elas estão disponíveis, elas contam na [avaliação da sua competência](#)?

Existem muitos vieses cognitivos em jogo: viés [retrospectivo](#), causalidade falsa, viés positivo, [ancoragem e pré-ativação](#). E, acima de tudo, o temido [viés de confirmação](#). Uma vez que uma ideia se instala em nossa mente, é provável ela permanecer lá.

### Referências

[1] Dale Griffin and Amos Tversky, “The Weighing of Evidence and the Determinants of Confidence,” *Cognitive Psychology* 24, no. 3 (1992): 411–435, doi:10.1016/0010-0285(92)90013-R.

---

41 NT.: Texto original em inglês. *Over the past few years, we have discreetly approached colleagues faced with a choice between job offers, and asked them to estimate the probability that they will choose one job over another. The average confidence in the predicted choice was a modest 66%, but only 1 of the 24 respondents chose the option to which he or she initially assigned a lower probability, yielding an overall accuracy rate of 96%.*

## 98 - Adiar propor soluções



De *Rational Choice in an Uncertain World* (Uma escolha racional em um mundo incerto), de Robyn Dawes [1]. O negrito adicionado é meu.

Norman R. F. Maier observou um padrão em grupos que enfrentam problemas. É comum que os membros proponham soluções enquanto ainda discutem o problema. Isso foca a interação nos méritos e problemas das propostas. As pessoas se apegam emocionalmente às suas ideias, e soluções melhores podem não surgir.

Para melhorar a solução de problemas em grupo, Maier propôs uma diretriz: **“Não proponha soluções até que o problema tenha sido discutido minuciosamente, sem sugerir nenhuma solução.”** Essa diretriz funciona bem quando existem boas soluções objetivamente definíveis.

Maier planejou um experimento de “interpretação de papéis” para demonstrar seu ponto. Ele reuniu três funcionários com habilidades diferentes em uma linha de montagem. Eles se revezavam em três empregos de diferentes níveis de habilidade. O funcionário mais capaz queria evitar o tédio. O menos capaz concordava com a rotação devido ao domínio do colega mais hábil.

Um “especialista em eficiência” observou que a produtividade poderia aumentar 20%. Sua recomendação: dar a tarefa mais difícil ao mais capaz e a mais fácil ao menos capaz. Os funcionários e um “capataz” foram convidados a discutir essa recomendação.

Alguns grupos receberam a diretriz de Maier, outros não. Os grupos sem a diretriz logo discutiam produtividade versus autonomia e tédio. Os grupos com a diretriz tinham mais chances de chegar a uma solução melhor: os dois funcionários mais aptos faziam rodízio, enquanto o menos apto ficava com a tarefa mais simples. Essa solução gerou um aumento de 19% na produtividade.

Tenho usado essa diretriz em muitos grupos que liderei, **especialmente com problemas difíceis. Nesses momentos, as pessoas tendem a propor soluções imediatamente.** Embora eu não tenha critérios objetivos, a diretriz de Maier parece promover melhores soluções<sup>42</sup>.

---

42 NT.: Texto original em inglês. Norman R. F. Maier noted that when a group faces a problem, the natural tendency of its members is to propose possible solutions as they begin to discuss the problem. Consequently, the group interaction focuses on the merits and problems of the proposed solutions, people become emotionally attached to the ones they have suggested, and superior solutions are not suggested. Maier enacted an edict to enhance group problem solving: “Do not propose solutions until the problem has been discussed as thoroughly as possible without suggesting any.” It is easy to show that this edict works in contexts where there are objectively defined good solutions to problems. Maier devised the following “role playing” experiment to demonstrate his point. Three employees of differing ability work on an assembly line. They rotate among three jobs that require different levels of ability, because the most able—who is also the most dominant—is strongly motivated to avoid boredom. In contrast, the least able worker, aware that he does not perform the more difficult jobs as well as the other two, has agreed to rotation because of the dominance of his able co-worker. An “efficiency expert” notes that if the most able employee were given the most difficult task and the least able the least difficult, productivity could be improved by 20%, and the expert recommends that the employees stop rotating. The three employees and... a fourth person designated to play the role of foreman are asked to discuss the expert’s recommendation. Some role-playing groups are given Maier’s edict not to discuss solutions until having discussed the problem thoroughly, while others are not. Those who are not given the edict immediately begin to argue about the

Isso é impressionantemente verdadeiro. E a situação piora com problemas mais complexos. Veja a Inteligência Artificial, por exemplo. É surpreendente quantas pessoas acham que sabem construir uma IA Geral. Isso sem conhecer problemas mais simples, como reconhecimento óptico de caracteres. E ao se tratar de uma IA com impacto positivo — uma [IA Amigável](#) — bem, esse problema é tão difícil que muitos o “resolvem” em quinze segundos.

Esse problema não é exclusivo da IA. Físicos encontram não físicos com teorias próprias sobre física. Economistas ouvem novas teorias surpreendentes sobre a economia. Se você é um biólogo evolutivo, qualquer pessoa pode “resolver” problemas em aberto no seu campo.

O conselho de Maier ecoa o princípio do [resultado final](#). A eficácia de nossas decisões depende apenas das evidências e do processamento que fazemos. Depois de escrever o resultado final, é [tarde para acrescentar razões](#). Se você decide muito cedo, sua decisão será baseada em pouca reflexão.

Além disso, [mudamos de ideia menos do que pensamos](#). Em um estudo, 24 pessoas atribuíram 66% de probabilidade à opção mais provável. Apenas 1 em 24 escolheu a opção menos provável. **Após adivinhar sua resposta, você provavelmente já decidiu.** Se você adivinha meio segundo após a pergunta, tem apenas meio segundo para ser inteligente.

A [Racionalidade Tradicional](#) enfatiza a falsificação: abandonar uma opinião inicial diante de evidências contrárias. Mas uma vez que uma ideia entra na sua cabeça, é difícil removê-la. Nem sempre temos evidências esmagadoras. Um método mais poderoso (e difícil) seria adiar o pensamento de uma resposta. Suspender aquele momento em que ainda não adivinhamos nossa resposta. Assim, damos mais tempo para nossa inteligência agir.

Mesmo meio minuto seria uma melhoria em relação a meio segundo.

## Referências

[1] Dawes, Rational Choice in An Uncertain World, 55–56.

---

*importance of productivity versus worker autonomy and the avoidance of boredom. Groups presented with the edict have a much higher probability of arriving at the solution that the two more able workers rotate, while the least able one sticks to the least demanding job—a solution that yields a 19% increase in productivity. I have often used this edict with groups I have led — particularly when they face a very tough problem, which is when group members are most apt to propose solutions immediately. While I have no objective criterion on which to judge the quality of the problem solving of the groups, Maier’s edict appears to foster better solutions to problems.*

## 99 - A falácia genética



Nas listas de [falácias lógicas](#), encontramos a “falácia genética”. É o erro de atacar uma crença com base em suas origens.

À primeira vista, isso pode parecer estranho. Se as causas de uma crença não determinam sua confiabilidade, o que determina? Quando o Deep Blue sugere uma jogada de xadrez, confiamos nele. Confiamos por entender o código que explora a árvore de jogadas. Não podemos avaliar a árvore por nós mesmos. O que justifica uma probabilidade como ‘racional’, senão sua origem em um processo confiável?

Os artigos sobre a falácia genética fazem uma distinção. O raciocínio genético nem sempre é falacioso. A origem da evidência pode ser relevante, como no caso de um especialista confiável. Porém, em outros casos, é uma falácia. Por exemplo, Kekulé viu a estrutura do benzeno em um sonho. Isso não significa que nunca possamos confiar nessa descoberta.

Então, às vezes a falácia genética é uma falácia e outras vezes não?

Formalmente, a falácia genética é considerada um erro. A causa original de uma crença não é o mesmo que seu status atual de justificação. O status atual considera todo o suporte e oposição conhecidos no momento.

Mesmo assim, [mudamos de ideia menos do que pensamos](#). As acusações genéticas têm um peso entre humanos que não teriam entre bayesianos ideais.

Limpar a mente é uma heurística poderosa. Usamos quando suspeitamos que muitas de nossas ideias venham de uma fonte defeituosa.

Uma vez que uma ideia se instala, nem sempre é fácil eliminá-la. Pense nas pessoas que cresceram acreditando na Bíblia. Mais tarde, podem rejeitar que Deus a escreveu. Mas ainda acreditam que ela contém sabedoria ética indispensável. Elas falharam em limpar suas mentes. Poderiam fazer melhor duvidando de tudo que a Bíblia diz, simplesmente porque ela diz.

Ao mesmo tempo, lembre-se: [estupidez invertida não é inteligência](#). O objetivo é agitar genuinamente sua mente e promover pensamento independente. Não é apenas rejeitar a Bíblia e deixá-la se tornar seu algoritmo.

Uma ideia em sua mente tende a encontrar apoio em todo lugar. Quando a fonte original é questionada, é sensato suspeitar de todas as conclusões derivadas.

Se for possível! Limpar sua mente não é fácil. É preciso esforço para reconsiderar, em vez de [ensaiar](#) argumentos [pré-armazenados](#). Como disse Thor Shenkel, “não é uma verdadeira crise de fé, a menos que as coisas possam acontecer facilmente de outra maneira”.

Desconfie se muitas ideias de uma fonte não confiável parecem corretas. A Bíblia é o exemplo óbvio.

Por outro lado, evidência clara torna a origem irrelevante. Acumular esse tipo de evidência é o objetivo da Ciência. Não importa mais que Kekulé tenha visto pela primeira vez a estrutura em anel do benzeno em um sonho — não importaria se tivéssemos encontrado a hipótese a ser testada gerando imagens aleatórias por computador, ou por um médium revelado como fraude, ou até mesmo na Bíblia. A estrutura em anel do benzeno está fundamentada em evidências experimentais suficientes para tornar irrelevante a origem da

sugestão.

Na ausência de evidências claras, preste atenção às fontes originais. Dê mais crédito a especialistas em áreas respeitadas. Desconfie de ideias de fontes duvidosas. Desconfie daqueles com motivos não confiáveis, se não apresentarem argumentos independentes.

A falácia genética é um erro quando há justificativas além do fato genético. O erro está em apresentar a acusação genética como se resolvesse toda a questão. Hal Finney [sugere](#) chamar o apelo correto à origem de “heurística genética”.

Algumas regras de ouro para humanos:

- Desconfie de acusações genéticas contra crenças que você não gosta. Especialmente quando há justificativas além da autoridade do orador. Exemplo: “O voo é uma ideia religiosa, então os irmãos Wright devem ser mentirosos”.
- Não pense que pode obter informações confiáveis apenas analisando personalidades e motivos. Se houver argumentos técnicos, estes devem ter prioridade.
- Quando suspeitar de uma fonte fundamental, duvide de tudo que se originou dela. Mas não descarte completamente as conclusões. Lembre-se: estupidez invertida não é inteligência.
- Fique extremamente desconfiado se ainda acreditar nas primeiras sugestões de uma fonte que rejeitou.



**J - Espirais de autodestruição**



## 100 - A heurística do afeto



A heurística do afeto ocorre quando impressões subjetivas de bom ou mau atuam como uma heurística, oferecendo julgamentos rápidos e intuitivos. Sentimentos agradáveis e desagradáveis têm um papel fundamental no raciocínio humano, e a heurística do afeto traz consigo alguns vieses adoráveis — alguns dos meus favoritos.

Vamos começar com um dos vieses relativamente menos surpreendentes. Imagine que você vai se mudar para uma nova cidade e precisa enviar um antigo relógio de pêndulo. No primeiro caso, o relógio pertencia ao seu avô e foi um presente especial dos seus avós no seu quinto aniversário. No segundo caso, o relógio foi um presente de um parente distante e você não tem um apego emocional especial a ele. Quanto você estaria disposto a pagar por um seguro que indenizasse US\$100 caso o relógio fosse perdido durante o transporte? De acordo com Hsee e Kunreuther, os participantes afirmaram estar dispostos a pagar mais que o dobro no primeiro caso [1]. Isso pode parecer racional — afinal, por que não pagar mais para proteger um objeto de maior valor? — até você perceber que o seguro não oferece proteção especial para o relógio, apenas paga a mesma quantia se qualquer um dos relógios for perdido. (E sim, foi esclarecido que o seguro era fornecido por uma empresa externa e não havia nenhum motivo especial para essa diferença.)

Bem, isso não parece muito absurdo. Talvez se possa argumentar que os participantes estavam buscando resultados emocionais, não financeiros — uma espécie de consolo na compra do seguro.

Agora, considere o seguinte: Yamagishi mostrou que os participantes consideravam uma doença mais perigosa quando era descrita como matando 1.286 pessoas a cada 10.000, em comparação com uma doença que tinha uma probabilidade de morte de 24,14% [2]. Aparentemente, a imagem mental de mil cadáveres é muito mais alarmante do que uma única pessoa com uma probabilidade maior de sobreviver do que de morrer.

Mas a situação piora ainda mais.

Considere que um aeroporto precisa decidir se deve gastar dinheiro para adquirir equipamentos novos, enquanto críticos argumentam que o dinheiro deveria ser investido em outros aspectos da segurança aeroportuária. Slovic et al. apresentaram a dois grupos de participantes argumentos a favor e contra a compra do equipamento, usando uma escala de resposta que variava de 0 (nenhum apoio) a 20 (apoio muito forte) [3]. Um grupo viu a medida descrita como capaz de salvar 150 vidas. O outro grupo viu a medida descrita como capaz de salvar 98% das 150 vidas. A hipótese que motivou o experimento foi a seguinte: salvar 150 vidas soaria vagamente bom — isso é muito? Pouco? — enquanto salvar 98% de algo é claramente ótimo, pois 98% está muito próximo do limite superior da escala de porcentagem. Resultado: o apoio médio para salvar 150 vidas foi de 10,4, enquanto o apoio médio para salvar 98% das 150 vidas foi de 13,6.

Outro exemplo interessante é o relatório de Denes-Raj e Epstein: Os participantes tiveram a oportunidade de ganhar US\$ 1 cada vez que retiravam aleatoriamente uma jujuba vermelha de uma tigela, preferindo tirar geralmente de uma tigela com mais jujubas vermelhas, mesmo que a proporção geral de jujubas vermelhas fosse menor. Por exemplo, eles preferiam uma chance de 7 em 100 a uma chance de 1 em 10.

De acordo com Denes-Raj e Epstein [4], esses participantes relataram posteriormente que, embora soubessem que as probabilidades estavam contra eles, sentiam terem uma chance melhor quando havia mais jujubas vermelhas. Isso pode parecer ilógico para um leitor familiarizado com estatística, mas se você refletir mais profundamente, perceberá que faz todo o sentido. Uma probabilidade de 7% contra 10% pode



ser uma má notícia, mas é mais do que compensada pelo aumento do número de jujubas vermelhas. É uma probabilidade pior, sim, mas ainda é mais provável que você ganhe, entende? Vale a pena refletir sobre esse pensamento até você alcançar a iluminação sobre como o resto do planeta pensa sobre probabilidade.

Finucane et al. descobriram que, para reatores nucleares, gás natural e conservantes de alimentos, apresentar informações sobre altos benefícios fazia com que as pessoas percebessem riscos menores; apresentar informações sobre maiores riscos fazia com que as pessoas percebessem benefícios menores; e assim por diante em outras situações [5]. As pessoas fundem seus julgamentos sobre aspectos positivos ou negativos particulares de algo em um sentimento geral bom, ou ruim sobre aquela coisa.

Finucane et al. também descobriram que a pressão do tempo aumentava consideravelmente a relação inversa entre o risco percebido e o benefício percebido, o que está de acordo com a descoberta geral de que a pressão do tempo, a falta de informação ou a distração aumentam a predominância da heurística afetiva sobre a deliberação analítica.

Ganzach encontrou o mesmo efeito no campo das finanças [6]. De acordo com a teoria econômica convencional, o retorno e o risco devem se correlacionar positivamente — em outras palavras, as pessoas pagam um preço premium por investimentos seguros, o que diminui o retorno; ações oferecem retornos mais altos do que títulos, mas também possuem um risco correspondente maior. Ao avaliar ações familiares, os analistas julgavam consistentemente os riscos e retornos positivamente correlacionados, como previsto. No entanto, ao avaliar ações desconhecidas, os analistas tendiam a julgá-las como geralmente boas ou ruins — baixo risco e alto retorno, ou alto risco e baixo retorno [7].

## Referências

- [1] Christopher K. Hsee and Howard C. Kunreuther, “The Affection Effect in Insurance Decisions,” *Journal of Risk and Uncertainty* 20 (2 2000): 141–159, doi:10.1023/A:1007876907268.
- [2] Kimihiko Yamagishi, “When a 12.86% Mortality Is More Dangerous than 24.14%: Implications for Risk Communication,” *Applied Cognitive Psychology* 11 (6 1997): 461–554.
- [3] Paul Slovic et al., “Rational Actors or Rational Fools: Implications of the Affect Heuristic for Behavioral Economics,” *Journal of Socio-Economics* 31, no. 4 (2002): 329–342, doi:10.1016/S1053-5357(02)00174-9.
- [4] Veronika Denes-Raj and Seymour Epstein, “Conflict between Intuitive and Rational Processing: When People Behave against Their Better Judgment,” *Journal of Personality and Social Psychology* 66 (5 1994): 819–829, doi:10.1037/0022-3514.66.5.819.
- [5] Finucane et al., “The Affect Heuristic in Judgments of Risks and Benefits.”
- [6] Yoav Ganzach, “Judging Risk and Return of Financial Assets,” *Organizational Behavior and Human Decision Processes* 83, no. 2 (2000): 353–370, doi:10.1006/obhd.2000.2914.
- [7] Slovic et al., “Rational Actors or Rational Fools.”

## 101 - Avaliabilidade (e compras de Natal baratas)



Com a temporada de [altos gastos de festas de fim de ano](#) se aproximando, uma pergunta pode estar surgindo na mente dos nossos ouvintes:

“Querido blog *Overcoming Bias*, existem formas de parecer generoso sem gastar muito dinheiro?”

Fico feliz em informar que sim! De acordo com um artigo de Hsee, intitulado “Menos é melhor: quando opções de baixo valor são mais valorizadas do que opções de alto valor”, presentear alguém com um cachecol de US\$ 45 pode te fazer parecer mais generoso do que dar um casaco de US\$ 55 [\[1\]](#).

Este é um caso específico de um fenômeno mais amplo. Em um experimento anterior, Hsee perguntou aos participantes quanto pagariam por um dicionário de música usado [\[2\]](#):

- Dicionário A, de 1993, com 10.000 verbetes, em perfeito estado.
- Dicionário B, de 1993, com 20.000 verbetes, com a capa rasgada, mas em perfeito estado por dentro.

Alguns participantes viram os dois dicionários lado a lado, enquanto outros viram apenas um. Quem viu apenas uma opção, em média, pagaria US\$24 pelo Dicionário A e US\$20 pelo B. Mas, quem viu os dois lado a lado, pagaria US\$27 pelo Dicionário B e US\$19 pelo A.

Claro, o número de verbetes é mais importante do que uma capa rasgada, se você for usar o dicionário. Mas, se você só tem um dicionário com 20.000 verbetes, esse número não significa muito sozinho. É um número alto? Considerável? Difícil dizer. Ele não é facilmente avaliável. Já a capa rasgada se destaca, por ter uma [valência afetiva](#) definida: é ruim.

No entanto, quando as duas opções são colocadas lado a lado, o número de verbetes se torna algo comparável, por haver duas quantidades para analisar. E, uma vez que o número de verbetes se torna avaliável, essa característica se sobressai em relação à capa rasgada.

Outro exemplo vem de Slovic e seus colegas [\[3\]](#): qual opção você prefere?

1. Vinte e nove chances em 36 de ganhar US\$2.
2. Sete chances em 36 de ganhar US\$9.

Embora os valores médios atribuídos a essas opções tenham sido US\$1,25 e US\$2,11, respectivamente, suas avaliações de atratividade foram 13,2 e 7,5. Tanto os preços quanto às avaliações de atratividade foram obtidos em um contexto em que os participantes foram informados de que duas apostas seriam selecionadas aleatoriamente entre as classificadas, e eles jogariam a aposta com o preço mais alto ou com a avaliação de atratividade mais alta. (Os participantes tinham um motivo para classificar as apostas como mais atraentes ou precificá-las mais alto, pois jogariam a aposta que preferissem.)

A aposta que oferecia um valor monetário maior parecia menos atraente, uma clássica inversão de preferência. Os pesquisadores levantaram a hipótese de que os valores em reais eram mais compatíveis com a tarefa de precificação, enquanto a probabilidade de retorno era mais compatível com a atratividade. Então, por que não tentar tornar o resultado da aposta mais emocionalmente relevante — mais afetivamente avaliável — e, assim, mais atraente?

E como eles fizeram isso? Adicionando uma pequena perda à aposta. A aposta original tinha 7 chanc-

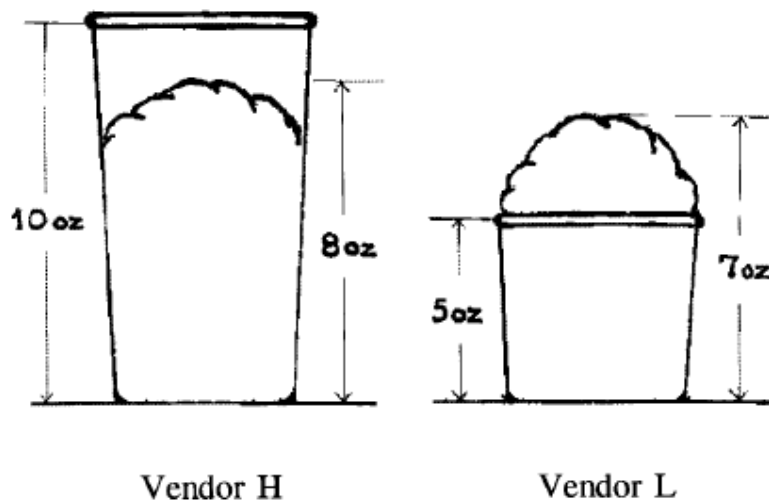
es em 36 de ganhar US\$9. A nova aposta tinha 7 chances em 36 de ganhar US\$9 e 29 chances em 36 de perder US\$0,05. Na aposta original, você avaliava a atratividade de ganhar US\$9. A nova aposta permitia que você avaliasse a atratividade de ganhar US\$9 contra perder US\$0,05.

“Os resultados”, afirmaram Slovic e seus colegas, “superaram nossas expectativas”. Em um novo experimento, a aposta simples com 7 chances em 36 de ganhar US\$9 teve uma média de classificação de atratividade de 9,4, enquanto a aposta complexa, que incluía a chance de perder US\$0,05, teve uma média de 14,9.

Um experimento de acompanhamento testou se os participantes preferiam a aposta original a um ganho certo de US\$2. Apenas 33% dos alunos preferiram a aposta original. Já entre outro grupo, que foi solicitado a escolher entre um ganho certo de US\$2 e a nova aposta (com a possibilidade adicional de uma perda de US\$0,05), 60,8% preferiram a aposta. Afinal, US\$9 não é uma quantia muito atraente, mas a relação entre US\$9 e US\$0,05 é extremamente atraente em termos de ganho/perda.

Você pode tornar uma aposta mais atraente adicionando uma perda mínima! A psicologia não é fascinante? É por isso que ninguém que verdadeiramente aprecie a maravilhosa complexidade da inteligência humana deseja projetar uma IA similar à humana. Claro, isso só funciona se os participantes não virem as duas apostas lado a lado.

Da mesma forma, qual desses dois sorvetes você acha que os participantes do estudo de Hsee de 1998 preferiram?



Naturalmente, a resposta depende se os participantes viram um único sorvete ou os dois, lado a lado. Os indivíduos que viram um único sorvete estavam dispostos a pagar US\$1,66 ao vendedor H e US\$2,26 ao vendedor L. Os indivíduos que viram os dois sorvetes estavam dispostos a pagar US\$1,85 ao vendedor H e US\$1,56 ao fornecedor L.

O que isso sugere para suas compras de fim de ano? Se você gastar US\$400 em um smartphone de última geração, seu destinatário verá um aparelho caro. Se você gastar US\$400 em um console de videogame, o destinatário verá o console mais básico. Qual é o melhor custo-benefício? Essa pergunta só faz sentido se você vir os dois produtos lado a lado. Durante suas compras, você pensará neles lado a lado, mas o destinatário verá apenas o que receber.

Se você tem uma quantia fixa para gastar — e seu objetivo é demonstrar amizade em vez de ajudar realmente o destinatário — será melhor não comprar com base no valor intrínseco. Decida quanto quer gastar para impressionar e encontre o objeto mais inútil que custe essa quantia. Quanto mais barata a categoria, mais valioso um objeto específico parecerá, desde que você gaste uma quantia fixa. O que é mais memorável: uma camisa de US\$25 ou uma vela de US\$25?

Isso dá um novo significado ao costume japonês de comprar melões por US\$50, não é mesmo? Você olha para isso, balança a cabeça e diz: “O que há com os japoneses?” E, no entanto, eles são vistos como incrivelmente generosos, mesmo gastando apenas US\$50. Você poderia gastar US\$200 em um jantar chique e não parecer tão generoso quanto gastar US\$50 em um melão. Se ao menos houvesse o costume de presentear palitos de dente por US\$25 ou grãos de poeira por US\$10; eles poderiam se safar gastando ainda menos.

PS: Se você realmente usar esse truque, gostaria de saber o que você comprou.

## Referências

- [1] Christopher K. Hsee, “Less Is Better: When Low-Value Options Are Valued More Highly than High-Value Options,” *Behavioral Decision Making* 11 (2 1998): 107–121.
- [2] Christopher K. Hsee, “The Evaluability Hypothesis: An Explanation for Preference Reversals between Joint and Separate Evaluations of Alternatives,” *Organizational Behavior and Human Decision Processes* 67 (3 1996): 247–257, doi:10.1006/obhd.1996.0077.
- [3] Slovic et al., “Rational Actors or Rational Fools.”

## 102 - Escalas ilimitadas, enormes prêmios do júri e futurismo



A psicofísica é o campo que estuda a relação entre os efeitos físicos e os efeitos sensoriais. Por exemplo, no caso da energia acústica, a pergunta é: quanta energia acústica a mais é necessária para um som parecer duas vezes mais alto para um ouvinte humano? Surpreendentemente, não é o dobro, mas sim cerca de oito vezes mais.

Medir energia acústica e fótons é fácil. Mas, para saber quão alto um som parece ou quão brilhante uma luz aparenta ser, geralmente precisamos perguntar ao ouvinte ou observador. Isso pode ser feito usando escalas limitadas, como de “muito quieto” a “muito alto” ou de “muito escuro” a “muito claro”.

Também podemos usar escalas ilimitadas, que começam em zero (nada audível ou visível) e aumentam indefinidamente. Nesses casos, o observador recebe um estímulo constante, chamado módulo, ao qual atribui um valor fixo. Por exemplo, um som recebe o valor 10. O observador pode então indicar que outro som é duas vezes mais alto, atribuindo o valor 20.

Essa abordagem tem se mostrado bastante confiável. Mas o que acontece se os participantes receberem uma escala ilimitada sem módulo de referência? Uma escala de zero ao infinito, sem um ponto fixo de comparação? Nesse caso, os participantes criam seu próprio módulo.

As proporções entre os estímulos permanecem consistentes entre os participantes. Por exemplo, se o participante A diz que o som X tem volume 10 e o som Y tem volume 15, e o participante B diz que o som X tem volume 100, é provável que ele atribua um volume próximo de 150 ao som Y. Porém, sem saber o módulo do participante C, seu fator de escala, não podemos prever o que ele dirá sobre o som X. Pode ser 1, pode ser 1.000.

Portanto, quando um participante avalia um estímulo em uma escala ilimitada, sem um padrão de comparação, quase toda a variação se deve à escolha arbitrária do módulo, e não ao estímulo em si.

Essa situação se assemelha à deliberação do júri em casos de danos punitivos. Não é surpreendente haver tanta variação nas decisões, pois cada jurado pode estar usando um módulo diferente ao atribuir valores em dólares aos danos.

Para demonstrar essa relação, Kahneman e seus colegas realizaram um estudo com 867 participantes elegíveis para júri. Eles apresentaram descrições de casos legais, como o de uma criança cujas roupas pegaram fogo, e pediram aos participantes que:

1. Classificassem a gravidade das ações do réu em uma escala limitada;
2. Classificassem o grau de punição que o réu deveria receber em uma escala limitada, ou;
3. Atribuísem um valor em dólares aos danos punitivos.[\[1\]](#)

No estudo, os participantes mostraram alta correlação em suas classificações de indignação e punição, mas os valores dos danos punitivos variaram muito. No entanto, a ordem das classificações dos valores — do menor para o maior — apresentou boa correlação entre os participantes.

Isso significa que, se você soubesse o cenário específico apresentado aos participantes — como o caso da criança — seria possível prever a classificação da punição e ter uma boa estimativa da classificação do valor em dólares em relação a outros casos. Porém, o valor exato em dólares seria imprevisível.

Para os valores brutos em dólares, a variação explicada foi de apenas 0,06 ! Tirar a média de doze respostas aleatórias também não ajudou muito.

Portanto, um valor de indenização por danos punitivos não é tanto uma avaliação econômica, mas sim uma expressão de atitude — uma medida psicofísica de indignação, manifesta em uma escala ilimitada sem ponto de referência fixo.

Observo que muitas previsões futuristas são, de maneira similar, melhor consideradas expressões de atitude. Faça a pergunta: “Quanto tempo levará até que tenhamos IA de nível humano?” As respostas que vi para essa pergunta estão espalhadas por todo o espectro. Em uma ocasião memorável, um indivíduo ligado à IA convencional me disse: “Quinhentos anos”. (!!)

Agora, a razão pela qual o tempo para alcançar a IA não é facilmente previsível é uma discussão longa por si só. Mas não é como se o indivíduo que mencionou “quinhentos anos” estivesse prevendo o futuro. E ele não poderia ter obtido esse número usando o método padrão equivocado da Lei de Moore. Então, o que ele quis dizer com o número 500?

Da forma como consigo imaginar, é como se eu tivesse perguntado: “Em uma escala em que zero é ‘nada difícil’, o quão difícil o problema da IA parece para você?” Se essa fosse uma escala limitada, todos os entrevistados marcariam “extremamente difícil” no extremo direito. Tudo parece extremamente difícil quando você não sabe como lidar com algo.

No entanto, nesse caso, temos uma escala ilimitada sem um ponto de referência fixo. Portanto, as pessoas simplesmente inventam um número para representar “extremamente difícil”, o que pode resultar em 50, 100 ou até mesmo 500. Em seguida, acrescentam “anos” no final, e essa é a sua previsão futurista.

“Quão difícil parece o problema da IA?” não é a única questão substituível. Outros responderiam como se eu tivesse perguntado: “Quão positivo você se sente em relação à IA?”, sendo que números mais baixos indicariam sentimentos mais positivos, e então acrescentam “anos” no final. No entanto, se essas “estimativas de tempo” representam algo além de expressões de atitude em uma escala ilimitada sem um ponto de referência fixo, não consegui determinar.

## Referências

[1] Daniel Kahneman, David A. Schkade, and Cass R. Sunstein, “Shared Outrage and Erratic Awards: The Psychology of Punitive Damages,” *Journal of Risk and Uncertainty* 16 (1 1998): 48–86, doi:10.1023/A:1007710408413; Daniel Kahneman, Ilana Ritov, and David Schkade, “Economic Preferences or Attitude Expressions?: An Analysis of Dollar Responses to Public Issues,” *Journal of Risk and Uncertainty* 19, nos. 1–3 (1999): 203–235, doi:10.1023/A:1007835629236.

## 103 - O efeito Halo



A heurística do afeto descreve como uma sensação geral de “bom” ou “mau” influencia muitos outros julgamentos. Isso acontece independentemente de serem lógicos ou não, e mesmo que você esteja ciente disso ou não.

Por exemplo, indivíduos informados sobre os benefícios da energia nuclear tendem a considerá-la menos arriscada. Analistas financeiros, ao avaliarem ações desconhecidas, as julgam como geralmente boas ou ruins. Baixo risco e alto retorno, ou alto risco e baixo retorno - desafiando a teoria econômica comum de que risco e retorno devem se correlacionar positivamente.

O efeito halo é a manifestação da heurística do afeto na psicologia social. Robert Cialdini, em seu livro *Influence: Science and Practice* (Influência: ciência e prática), resume [\[1\]](#):

Pesquisas mostraram que atribuímos automaticamente a pessoas bonitas características favoráveis, como talento, gentileza, honestidade e inteligência (para revisão desta evidência, consulte Eagly, Ashmore, Makhijani e Longo, 1991) [\[2\]](#). Além disso, fazemos esses julgamentos sem estar cientes de que a atratividade física desempenha um papel nesse processo.

Algumas consequências dessa suposição inconsciente de que “bonito é igual a bom” são preocupantes. Um estudo das eleições federais no Canadá em 1974 descobriu que candidatos atraentes receberam mais que o dobro dos votos em comparação com candidatos menos atraentes (Efran and Patterson, 1976) [\[3\]](#). Apesar dessas evidências de favoritismo em relação a políticos bonitos, as pesquisas subsequentes mostraram que os eleitores não perceberam seu preconceito. Na verdade, 73% dos eleitores canadenses pesquisados negaram veementemente que a aparência física tenha influenciado seus votos, e apenas 14% admitiram a possibilidade de tal influência (Efran and Patterson, 1976) [\[4\]](#). Os eleitores podem negar o quanto quiserem o impacto da atratividade na elegibilidade, mas as evidências continuam confirmando sua preocupante presença (Budesheim and DePaola, 1994) [\[5\]](#).

Um efeito semelhante foi observado em situações de contratação. Em um estudo, a aparência física dos candidatos em uma simulação de entrevista de emprego teve um papel mais significativo nas decisões de contratação do que as qualificações para o trabalho — isso apesar dos entrevistadores afirmarem que a aparência desempenhava um papel menor em suas escolhas (Mack and Rainey, 1990) [\[6\]](#). A vantagem dada aos trabalhadores atraentes se estende desde o momento da contratação até o pagamento. Economistas que examinaram amostras nos Estados Unidos e no Canadá descobriram que indivíduos atraentes recebem em média de 12% a 14% mais do que seus colegas de trabalho menos atraentes (Hamermesh and Biddle, 1994) [\[7\]](#).

Pesquisas igualmente preocupantes indicam que nosso sistema judiciário também é suscetível à influência de características físicas, como aparência e estrutura óssea. Parece que pessoas bonitas provavelmente receberão tratamento altamente favorável no sistema jurídico (consulte Castellow, Wuensch, e Moore, 1991; e Downs e Lyons, 1990, para revisões) [\[8\]](#). Por exemplo, em um estudo realizado na Pensilvânia (Stewart, 1980) [\[9\]](#), pesquisadores avaliaram a atratividade física de 74 réus do sexo masculino no início de seus julgamentos criminais.

Quando, mais tarde, os pesquisadores verificaram os registros do tribunal para analisar os

resultados desses casos, descobriram que os réus atraentes receberam sentenças significativamente mais leves. De fato, os réus atraentes tinham duas vezes mais chances de evitar a prisão em comparação com os réus menos atraentes. Em outro estudo sobre indenizações concedidas em um julgamento por negligência, quando o réu era mais atraente do que a vítima, a quantia média concedida era de US\$5.623, mas quando a vítima era a mais atraente, a compensação média era de US\$10.051. Além disso, tanto jurados do sexo masculino quanto do sexo feminino mostraram favoritismo com base na atratividade (Kulka and Kessler, 1978) [\[10\]](#).

Outros experimentos demonstram que pessoas atraentes têm maior probabilidade de receber ajuda quando precisam (Benson, Karabenic, and Lerner, 1976) [\[11\]](#) e são mais persuasivas na mudança de opiniões do público (Chaiken, 1979)...[\[12\]](#)

A influência da atratividade nas avaliações de inteligência, honestidade ou bondade é um exemplo claro de viés, especialmente quando essas qualidades são julgadas com base em um texto fixo. Não esperaríamos que os julgamentos de honestidade e atratividade se misturassem por qualquer motivo legítimo.

Por outro lado, quanta da minha percepção de inteligência é atribuída à minha honestidade? Quanta da minha honestidade percebida é devida à minha inteligência? Encontrar a verdade e falar a verdade não estão tão amplamente separados na natureza quanto parecer bonito e parecer inteligente.

No entanto, esses estudos sobre o efeito halo da atratividade nos levam a suspeitar que possa haver um efeito halo semelhante para bondade ou inteligência. Suponha que você conheça alguém que não apenas pareça muito inteligente, mas também honesto, altruísta, gentil e sereno. É razoável suspeitar que algumas dessas características percebidas estão influenciando sua percepção das outras. Talvez a pessoa seja genuinamente inteligente, honesta e altruísta, mas não tão gentil ou serena. Você deve estar ciente de que as pessoas que você conhece podem não se dividir tão claramente entre demônios e anjos.

E — embora eu saiba que você não acredita que precise fazer isso, talvez deva — seja um pouco mais cético em relação aos candidatos políticos mais atraentes.



## Referências

- [1] Robert B. Cialdini, *Influence: Science and Practice* (Boston: Allyn & Bacon, 2001).
- [2] Alice H. Eagly et al., "What Is Beautiful Is Good, But . A Meta-analytic Review of Research on the Physical Attractiveness Stereotype," *Psychological Bulletin* 110 (1 1991): 109–128, doi:10.1037/0033-2909.110.1.109.
- [3] M. G. Efran and E. W. J. Patterson, "The Politics of Appearance" (Unpublished PhD thesis, 1976).
- [4] Ibid.
- [5] Thomas Lee Budesheim and Stephen DePaola, "Beauty or the Beast?: The Effects of Appearance, Personality, and Issue Information on Evaluations of Political Candidates," *Personality and Social Psychology Bulletin* 20 (4 1994): 339–348, doi:10.1177/0146167294204001.
- [6] Denise Mack and David Rainey, "Female Applicants' Grooming and Personnel Selection," *Journal of Social Behavior and Personality* 5 (5 1990): 399–407.
- [7] Daniel S. Hamermesh and Jeff E. Biddle, "Beauty and the Labor Market," *The American Economic Review* 84 (5 1994): 1174–1194.
- [8] Wilbur A. Castellow, Karl L. Wuensch, and Charles H. Moore, "Effects of Physical Attractiveness of the Plaintiff and Defendant in Sexual Harassment Judgments," *Journal of Social Behavior and Personality* 5 (6 1990): 547–562; A. Chris Downs and Phillip M. Lyons, "Natural Observations of the Links Between Attractiveness and Initial Legal Judgments," *Personality and Social Psychology Bulletin* 17 (5 1991): 541–547, doi:10.1177/0146167291175009.
- [9] John E. Stewart, "Defendants' Attractiveness as a Factor in the Outcome of Trials: An Observational Study," *Journal of Applied Social Psychology* 10 (4 1980): 348–361, doi:10.1111/j.1559-1816.1980.tb00715.x.
- [10] Richard A. Kulka and Joan B. Kessler, "Is Justice Really Blind?: The Effect of Litigant Physical Attractiveness on Judicial Judgment," *Journal of Applied Social Psychology* 8 (4 1978): 366–381, doi:10.1111/j.1559-1816.1978.tb00790.x.
- [11] Peter L. Benson, Stuart A. Karabenick, and Richard M. Lerner, "Pretty Pleases: The Effects of Physical Attractiveness, Race, and Sex on Receiving Help," *Journal of Experimental Social Psychology* 12 (5 1976): 409–415, doi:10.1016/0022-1031(76)90073-1.
- [12] Shelly Chaiken, "Communicator Physical Attractiveness and Persuasion," *Journal of Personality and Social Psychology* 37 (8 1979): 1387–1397, doi:10.1037/0022-3514.37.8.1387.

## 104 - Viés de super-herói



Suponhamos que haja um sociopata armado, um sequestrador com reféns, que rejeitou todas as tentativas de negociação e anunciou sua intenção de matar. Na vida real, os mocinhos geralmente não arrombam a porta quando o bandido tem reféns. Mas às vezes — raramente, mas às vezes — a vida imita Hollywood, e mocinhos genuínos precisam arrombar a porta.

Imagine, em duas realidades distintas, dois heróis invadindo a sala para enfrentar o vilão. Em uma realidade, o herói é forte o suficiente para lançar carros, disparar rajadas de energia e sua pele desvia balas. O vilão se escondeu em uma escola, com mais de duzentas crianças como reféns; seus pais estão do lado de fora, desesperados.

Na outra realidade, o herói é um policial e os reféns são três pessoas que o vilão capturou na rua. Reflita: quem é o maior herói? Quem tem mais chances de ter sua própria história em quadrinhos?

O [efeito halo](#) é a tendência de percebermos todos os traços positivos como correlacionados. Perfis altamente classificados em atratividade também recebem altas classificações em talento, bondade, honestidade e inteligência.

Portanto, personagens de quadrinhos que parecem fortes e invulneráveis também parecem ter mais traços heroicos de coragem. Como Adam Warren coloca em *Empowered* (Empoderado):

Quão difícil pode ser agir com bravura quando você é praticamente invulnerável?<sup>43</sup>

— Adam Warren, *Empowered* (Empoderado), Vol.1 [\[1\]](#)

Não consigo lembrar se li isso em algum lugar ou se criei essa hipótese: a fama, em particular, parece se combinar de forma aditiva com todas as outras características de personalidade.

Consideremos Gandhi. Gandhi foi a pessoa mais altruísta do século XX ou apenas o altruísta mais famoso? Gandhi confrontou a polícia com cassetetes e soldados armados. Porém, Gandhi era uma celebridade e desfrutava da proteção que a fama lhe proporcionava. E quanto aos outros manifestantes, as pessoas que enfrentaram cassetetes e armas, mesmo sem se tornarem manchetes internacionais, seriam mencionadas se fossem detidas ou baleadas?

O que Gandhi pensou ao buscar as manchetes, a celebridade, a fama, o lugar na história, ao se tornar o arquétipo da resistência não violenta, quando ele corria menos riscos do que qualquer uma das pessoas que marchavam ao seu lado? Como ele se sentiu quando um desses heróis anônimos se aproximou dele, com os olhos brilhando, e disse o quanto ele era maravilhoso? Gandhi teria visualizado seu mundo sob essas perspectivas? Não sei; afinal, não sou Gandhi.

Isso não é, de forma alguma, uma crítica a Gandhi. O objetivo da resistência não violenta é não exibir coragem. Isso pode ser feito com muito mais facilidade ao atravessar as Cataratas do Niágara em um barril. Gandhi não podia deixar de estar um pouco, embora não completamente, protegido por sua fama. E as ações

---

43 NT.: Texto original em inglês. *How tough can it be to act all brave and courageous when you're pretty much invulnerable?*

de Gandhi requeriam coragem — não tanta coragem quanto marchar anonimamente, mas ainda assim coragem.

O viés que gostaria de destacar é que a pontuação de fama de Gandhi parece ser perceptivelmente acrescentada à sua pontuação de altruísmo legítimo. Quando se pensa em não violência, pensa-se em Gandhi — não em um manifestante anônimo em uma das marchas de Gandhi que enfrentou grupos de choque e armas, foi espancado, teve de ser levado ao hospital e mancou pelo resto de sua vida, e cujo nome nunca foi lembrado.

Da mesma forma, o que é maior: arriscar a vida para salvar duzentas crianças ou arriscar a vida para salvar três adultos?

A resposta depende do que você quer dizer com “maior”. Se tiver que escolher entre salvar duzentas crianças e salvar três adultos, escolha a primeira opção. “Quem salva uma vida, salva o mundo inteiro” pode ser um belo sinal de aplausos, mas é um conselho terrível quando se precisa escolher entre um ou outro. Portanto, se você se refere a “maior” no sentido de “O que é mais importante?” ou “Qual resultado é preferível?”, então é maior salvar duzentos em vez de três.

Mas se você está perguntando sobre grandeza no sentido de virtude revelada, então alguém que arriscaria sua vida para salvar apenas três vidas revela mais coragem do que alguém que arriscaria sua vida para salvar duzentas, mas não três.

Isso não significa que você possa deliberadamente escolher arriscar sua vida para salvar três adultos e deixar as duzentas crianças à própria sorte, simplesmente porque deseja demonstrar mais virtude. Aquele que arrisca sua vida com o intuito de ser virtuoso revela muito menos virtude do que aquele que arrisca sua vida para salvar os outros. Aquele que escolhe salvar três vidas em vez de duzentas, acreditando que isso demonstra uma virtude maior, está tão obcecado com sua própria “grandeza” que comete um equivalente moral a um homicídio culposo.

É um daqueles cenários de *wu wei*<sup>44</sup>: você não pode revelar a virtude ao tentar revelar a virtude. Diante da escolha entre um método seguro para salvar o mundo que não envolve nenhum sacrifício pessoal ou desconforto, e um método que coloca sua vida em risco e exige grandes privações, não é possível se tornar um herói ao escolher deliberadamente o segundo caminho. Não há nada de heroico em desejar parecer um herói. Seria um propósito fadado ao fracasso.

As pessoas verdadeiramente virtuosas, tentam genuinamente salvar vidas em vez de buscar a revelação da virtude, sempre buscarão salvar mais vidas com menos esforço, o que significa que menos de sua virtude será revelada. Pode parecer confuso, mas não é contraditório.

No entanto, nem sempre conseguimos escolher ser invulneráveis a balas. Após fazermos o possível para reduzir os riscos e ampliar nosso alcance, qualquer heroísmo restante é verdadeiramente revelado.

O policial que coloca sua vida em risco sem superpoderes, sem visão de raio-x, sem superforça, sem a habilidade de voar e, acima de tudo, sem invulnerabilidade a balas, revela uma virtude muito maior do que o Super-Homem, que é apenas um super-herói comum.

## Referências

[1] Adam Warren, *Empowered*, vol. 1 (Dark Horse Books, 2007).

---

44 NT: *Wu wei* (無為), um conceito central do Taoísmo, significa literalmente “não-agir” ou “ação sem esforço”. Originário dos caracteres chineses “wu” (無), que denota ausência ou não fazer, e “wei” (為), que significa ação ou fazer, o termo paradoxal encapsula a ideia de agir em harmonia com o fluxo natural do Tao, alcançando resultados sem esforço ou força excessiva.

## 105 - Meros Messias



Já abordei o [efeito halo](#), que leva as pessoas a verem todas as características positivas como correlacionadas. Pessoas mais atraentes, por exemplo, também são vistas como mais gentis, honestas e inteligentes. Isso nos leva a admirar mais os heróis super fortes e imunes a balas. Mas, logicamente, é preciso muito mais coragem para ser um herói quando se é vulnerável. Além disso, agir corajosamente para salvar uma vida revela mais virtude do que salvar o mundo inteiro. (Embora, se você tiver que escolher, salve o mundo, é claro!)

Mas vamos a um exemplo mais concreto.

John Perry era um policial de Nova York, extropiano e transumanista — foi assim que soube dele. Perry estava prestes a se aposentar e abrir seu próprio escritório de advocacia, quando um avião atingiu o World Trade Center. Ele morreu quando a torre norte desabou. Não o conheci pessoalmente, então não posso afirmar com certeza, mas a maioria dos extropianos não acredita em Deus, e imagino que Perry também fosse ateu.

Ou seja, Perry sabia que arriscava a própria vida toda semana no trabalho. E não era como a maioria das pessoas na história, que sabiam que só podiam escolher como morrer e decidiram fazer valer a pena. Perry era transumanista; ele tinha esperança genuína. Mesmo assim, arriscou a vida. Não por esperar recompensa divina, nem por esperar algo após a morte. Mas, porque havia outras pessoas em perigo, que também não tinham almas imortais, e sua vida não valia mais do que a delas.

Não conheci John Perry. Não sei se ele via o mundo assim. Mas o fato de um ateu e transumanista poder ser policial, poder correr para um prédio em chamas, diz mais sobre o espírito humano do que todos os mártires que já almejaram o paraíso.

Este é um policial em particular...

... e agora, o super-herói.

Segundo os cristãos, Jesus Cristo andava sobre as águas, acalmava tempestades e expulsava demônios. Que vida confortável! Fome? Multiplique pães. Não gosta da figueira? Amaldiçoe-a. Problemas com os romanos? Deixe com seu Pai. Mas essa vida encantada acabou quando Jesus se entregou para a crucificação. Ser pregado em uma cruz não é uma morte agradável. Contudo, segundo os cristãos, Jesus sabia que ressuscitaria após três dias e iria para o céu. Qual ameaça o levou a enfrentar esse sofrimento temporário, seguido da eternidade no paraíso? A vida de uma pessoa? A corrupção da igreja na Judéia ou a opressão romana? Não. Segundo os cristãos, o destino eterno de toda a humanidade estava em jogo.

Mas não quero condenar alguém que não seja verdadeiramente culpado. E se Jesus — ou melhor, vamos usar seu nome correto: *Yeishu*<sup>45</sup> — E se *Yeishu* de Nazaré nunca tivesse andado sobre as águas, mas ainda assim tivesse desafiado a igreja da Judéia, estabelecida pelo poder romano? Isso não mereceria mais honra do que a dada a Jesus Cristo, que era apenas um messias comum?

Infelizmente, parece mais admirável um herói com superpoderes. Parece revelar mais virtude morrer

---

45 NT.: *Yeishu* (יֵשׁוּ) é uma forma abreviada e menos comum do nome *Yeishua* em hebraico. A palavra hebraica *Yeshua* (יְשׁוּעָה) é a forma original do nome “Jesus” em hebraico. Significa “Yahweh é salvação” ou “Yahweh salva”. É importante notar que, embora “Yeshua” seja a forma original do nome “Jesus” em hebraico, *Yeishu* carrega conotações negativas e é frequentemente usado em contextos pejorativos ou polêmicos, especialmente em algumas literaturas rabínicas.

temporariamente para salvar o mundo do que morrer de vez, confrontando uma igreja corrupta. Parece tão comum, como se muitos outros na história tivessem feito o mesmo.

Confortavelmente sentados dois mil anos no futuro, podemos criticar *Yeishu*, mas ele fez o que acreditava ser certo: confrontou uma igreja que considerava corrupta e morreu por isso. Sem o [benefício do tempo](#), dificilmente poderia prever o impacto real de sua vida no mundo. Comparado a outros profetas da época, ele provavelmente era mais honesto, menos violento e mais corajoso. Se ignorarmos as consequências não intencionais, o pior que se pode dizer de *Yeishu* é que outros na história se saíram melhor (Epicuro, Buda e Marco Aurélio vêm à mente). *Yeishu* morreu para sempre e, de certa forma, o fez por honestidade. Mil e quinhentos anos antes da ciência, a honestidade religiosa não era um paradoxo.

Como disse Sam Harris [\[1\]](#):

Não basta que Jesus tenha sido um homem que se transformou a tal ponto que o Sermão da Montanha pudesse ser a confissão de seu coração. Ele também precisava ser o Filho de Deus, nascido de uma virgem e destinado a retornar à Terra em glória. O efeito desse dogma é colocar o exemplo de Jesus permanentemente fora de nosso alcance. Seu ensinamento deixa de ser um conjunto de afirmações empíricas sobre a conexão entre ética e percepção espiritual e, em vez disso, se torna um conto de fadas gratuito e bastante terrível. Segundo o dogma do cristianismo, tornar-se como Jesus é impossível. Só se pode enumerar os próprios pecados, acreditar no inacreditável e aguardar o fim do mundo<sup>46</sup>.

Duvido muito que *Yeishu* tenha realmente proferido o Sermão da Montanha. Todavia, *Yeishu* merece honra. Ele merece mais honra do que os cristãos lhe concedem.

Mas, como *Yeishu* provavelmente acreditava que sua alma sobreviveria, ele não merece mais honra do que John Perry.

## Referências

[1] Sam Harris, *The End of Faith: Religion, Terror, and the Future of Reason* (WW Norton & Company, 2005).

---

46 NT.: Texto original em inglês. *It is not enough that Jesus was a man who transformed himself to such a degree that the Sermon on the Mount could be his heart's confession. He also had to be the Son of God, born of a virgin, and destined to return to earth trailing clouds of glory. The effect of such dogma is to place the example of Jesus forever out of reach. His teaching ceases to become a set of empirical claims about the linkage between ethics and spiritual insight and instead becomes a gratuitous, and rather gruesome, fairy tale. According to the dogma of Christianity, becoming just like Jesus is impossible. One can only enumerate one's sins, believe the unbelievable, and await the end of the world.*

## 106 - Espirais de autodestruição afetiva



São [muitas](#), muitas, [muitas](#) as falhas no raciocínio humano que nos levam a superestimar o quão bem nossa teoria favorita explica os fatos. A teoria química do flogisto poderia explicar praticamente qualquer coisa, desde que não tivesse que prever com antecedência. Quanto mais fenômenos você usa sua teoria favorita para explicar, mais verdadeira ela parece — afinal, não foi confirmada por todas essas observações? À medida que a teoria parece mais verdadeira, você fica mais inclinado a questionar evidências que entram em conflito com ela. E como a teoria preferida parece mais abrangente, você tentará usá-la para explicar cada vez mais coisas.

Se você conhece alguém que acredita que a Bélgica controla secretamente o sistema bancário dos EUA, ou que possa usar uma força invisível de espíritos azuis para detectar vagas de estacionamento disponíveis, é provável ter começado assim.

(Apenas observe atentamente e você verá muitas coisas que parecem confirmar essa teoria...)

Esse ciclo de feedback positivo de credulidade e confirmação é realmente assustador e responsável por muitos erros, tanto na ciência quanto na vida cotidiana.

No entanto, isso não se compara à espiral descendente que começa com uma carga de afeto positivo — um pensamento que faz você se sentir muito bem.

Um novo sistema político que pode salvar o mundo. Um líder grandioso, forte, nobre e sábio. Um tônico incrível que pode curar dores de estômago e câncer.

Por que não ter tudo isso? Uma grande causa precisa de um grande líder. Um grande líder deve poder criar um ou dois tônicos mágicos.

O [efeito halo](#) é que qualquer característica positiva percebida (como atração ou força) aumenta a percepção de qualquer outra característica positiva (como inteligência ou coragem). Mesmo quando faz [pouco ou nenhum sentido](#).

As características positivas aumentam a percepção de todas as outras características positivas? Isso parece muito com a fissão de um átomo de urânio, onde os nêutrons liberados dividem outros átomos de urânio.

O afeto positivo fraco é subcrítico, não sai do controle. Uma pessoa atraente parece mais honesta, o que talvez a torne ainda mais atraente, mas o fator efetivo de multiplicação dos nêutrons é inferior a um. Metaforicamente falando. A ressonância confunde um pouco as coisas, mas depois desaparece.

Com um intenso efeito positivo ligado à Grande Coisa, a ressonância se espalha por todos os lugares. Um comunista fervoroso vê a sabedoria de Marx em cada hambúrguer comprado no McDonald's, em todas as promoções negadas que deveriam ter sido suas em um verdadeiro paraíso do trabalhador, em todas as eleições que não lhe agradam, em cada artigo de jornal "inclinado na direção errada". Cada vez que eles usam a Grande Ideia para interpretar outro evento, a Grande Ideia é cada vez mais confirmada. Parece melhor - reforço positivo - e, é claro, quando algo parece bom, infelizmente, nos faz querer acreditar ainda mais.

Quando a Grande Coisa nos faz sentir bem o suficiente para buscar constantemente novas oportunidades para nos sentirmos ainda melhor sobre ela, aplicando-a para interpretar novos acontecimentos todos

os dias, a ressonância do afeto positivo se torna como [uma sala cheia de armadilhas de rato carregadas com bolas de pingue-pongue](#).

Você poderia chamar isso de um “atrator de felicidade”, “feedback excessivamente positivo”, um “ciclo vicioso de elogios” ou “ilusão de ótica”. Pessoalmente, prefiro o termo “espiral da morte afetiva”.

A seguir: como resistir a uma espiral da morte afetiva. (Dica: não é recusar-se a admirar qualquer coisa novamente, nem manter as coisas que você admira sob um controle estrito e seguro.)

## 107 - Resista à espiral da autodestruição feliz



Era uma vez um homem que estava convencido de possuir uma Grande Ideia. Na verdade, quanto mais ele mergulhava nessa Grande Ideia, mais ele percebia que não era apenas uma ideia grandiosa, mas a mais maravilhosa de todas. Essa Grande Ideia revelaria os mistérios do universo, desafiaria a autoridade corrupta e falha da Ordem Social Estabelecida, concederia poderes quase mágicos aos seus detentores, alimentaria os famintos, curaria os doentes e tornaria o mundo um lugar melhor, entre outros benefícios.

Esse homem era Francis Bacon, e sua Grande Ideia era o método científico. Ele foi o único visionário na história a reivindicar tamanhos benefícios para a humanidade e estar completamente certo.

(É claro que Bacon não inventou a ciência sozinho, mas ele contribuiu e pode ter sido o primeiro a reconhecer o seu poder.)

Eis o problema de decidir nunca admirar tanto assim: algumas ideias são realmente boas. Embora ninguém tenha feito reivindicações mais audaciosas do que Bacon, pelo menos até agora.

Mas como resistir à [espiral da autodestruição feliz](#) em relação à própria ciência? A espiral da autodestruição feliz começa quando você acredita que algo é tão maravilhoso que o [efeito halo](#) faz com que você encontre cada vez mais coisas boas sobre isso, fazendo-o parecer ainda mais maravilhoso, e assim por diante, em uma espiral descendente. E se a ciência for realmente tão benéfica que não possamos reconhecer sua verdadeira grandeza e manter nossa sanidade? Parece algo interessante de se dizer, não é? Oh, não, está começando! Corraaaa...

Se você recuperar a [sabedoria profunda padrão](#) para [não exagerar na admiração pela ciência](#), encontrará pensamentos como “A ciência nos deu o ar-condicionado, mas também criou a bomba de hidrogênio” ou “A ciência pode nos falar sobre estrelas e biologia, mas nunca pode provar ou refutar o dragão na minha garagem”. No entanto, as pessoas que originaram esses pensamentos não estavam tentando resistir a uma espiral da autodestruição feliz. Elas não estavam preocupadas com sua própria admiração pela ciência saindo do controle. Provavelmente, elas não gostaram do que a ciência tinha a dizer sobre suas crenças favoritas e procuraram maneiras de minar sua autoridade.

As críticas comuns à ciência provavelmente não atrairão aqueles que verdadeiramente se exaltam com ela — esse não é o público-alvo. Portanto, precisaremos encontrar outras perspectivas negativas para abordar.

Mas, ao procurar seletivamente algo negativo para dizer sobre a ciência — mesmo na tentativa de resistir a uma espiral de exaltação sem limites — não estaríamos nos condenando automaticamente à [racionalização](#)? Por que prestarmos atenção em nossos próprios pensamentos se sabemos que estamos tentando nos [manipular](#)?

Sou geralmente cético em relação às pessoas que afirmam que um viés pode ser usado para neutralizar outro. Isso me parece como um mecânico de automóveis que diz que o motor do limpador de para-brisa direito está quebrado, mas, ao invés de consertá-lo, ele simplesmente quebra o limpador do para-brisa esquerdo para equilibrar as coisas. Esse tipo de astúcia leva a um tiro no próprio pé. Seja qual for a solução, ela deve envolver acreditar em coisas verdadeiras, em vez de acreditar que acreditamos em coisas que sabemos serem falsas.

Será possível evitar essa espiral de exaltação sem limites restringindo nossa admiração pela ciência



a um domínio mais específico? Parte dessa espiral envolve enxergar a Grande Ideia em todos os lugares — como imaginar como o comunismo poderia curar o câncer se tivesse uma chance. Provavelmente, o sinal mais evidente de um guru de seita é quando ele alega ter experiência, não em apenas uma área, nem mesmo em um grupo de áreas relacionadas, mas em tudo. O guru sabe o que os membros da seita devem comer, vestir, fazer para ganhar a vida, com quem devem se relacionar, a qual forma de arte devem apreciar, que música devem ouvir...

Infelizmente, a maioria das pessoas falha miseravelmente quando tenta delimitar a pequena caixa na qual a ciência deve se encaixar. O truque comum de dizer “Ei, a ciência não irá curar o câncer” não funciona. “A ciência não tem nada a dizer sobre o amor de um pai por seu filho” — desculpe, mas isso é simplesmente falso. Se tentarmos separar a ciência do amor parental, não apenas estamos negando a ciência cognitiva e a psicologia evolutiva, mas também estamos negando a fundação da *United Therapeutics* por Martine Rothblatt, que buscou a cura para a hipertensão pulmonar de sua filha (e obteve sucesso, devo acrescentar). A ciência está legitimamente ligada, garantidamente, a quase todos os aspectos importantes da existência humana.

Então, qual seria um exemplo de uma afirmação falsa e válida que poderíamos fazer sobre a ciência?

Na minha humilde opinião, uma afirmação falsa seria dizer que a ciência é tão maravilhosa que [os cientistas nem mesmo deveriam tentar assumir a responsabilidade ética por seu trabalho](#), pois isso automaticamente resultaria em um final positivo. Essa afirmação, para mim, parece não compreender a natureza do processo pelo qual a ciência beneficia a humanidade. Os cientistas são seres humanos, possuem preocupações pró-sociais como a maioria das outras pessoas, e é pelo menos parte do motivo pelo qual a ciência acaba fazendo mais bem do que mal.

Mas esse ponto, evidentemente, não é indiscutível. Portanto, aqui está uma afirmação falsa e mais simples: “Um paciente com câncer pode ser curado apenas publicando artigos suficientes em periódicos”. Ou, “Os sociopatas podem se tornar completamente normais se comprometerem a nunca acreditar em nada sem evidências experimentais replicadas com  $p < 0,05$ ”.

A maneira de evitar acreditar em tais declarações não é limitar nossos afetos, decidindo que a ciência é apenas um pouco boa. Também não é procurar razões para acreditar que a publicação de artigos em periódicos causa câncer. Nem é acreditar que a ciência não tem nada a dizer sobre o câncer de forma alguma.

Em vez disso, se você compreende com bastante [especificidade](#) como a ciência funciona, então você sabe que, embora seja possível para a “ciência curar o câncer”, um paciente com câncer que escreve artigos para jornais não experimentará uma remissão milagrosa. Essa sequência proposta específica de causa e efeito não funcionará.

A espiral da autodestruição feliz é apenas um problema emocional decorrente de um problema de percepção, o efeito halo, que nos torna mais propensos a aceitar reivindicações positivas futuras uma vez que aceitamos uma afirmação positiva inicial. Não podemos nos livrar desse efeito apenas desejando; provavelmente, sempre nos influenciará um pouco. No entanto, podemos desacelerar, parar e considerar cada reivindicação positiva adicional como um detalhe adicional oneroso e nos concentrar nos pontos específicos da reivindicação além de sua positividade.

E se uma afirmação específica “não puder ser refutada”, mas houver argumentos “tanto a favor quanto contra”? Na verdade, essas são palavras que devemos abordar com cautela, porque muitas vezes é isso que as pessoas dizem quando estão [ensaiando as evidências](#) ou [evitando os verdadeiros pontos fracos](#). Dado o perigo da espiral da autodestruição feliz, faz sentido tentar evitar ficar satisfeitos com reivindicações não resolvidas — evitando transformá-las em uma fonte de ainda mais positividade sobre algo que já gostamos.

A espiral da autodestruição feliz é apenas um grande problema emocional devido ao feedback excessivamente positivo, a capacidade do processo de se tornar crítico. Você pode não conseguir eliminar completamente o efeito halo, mas pode aplicar raciocínio crítico suficiente para manter os halos subcríticos — garantindo que a ressonância desapareça em vez de explodir.

Podemos até dizer que todo o problema começa com as pessoas que não se preocupam em examinar criticamente cada detalhe adicional oneroso — exigindo evidências suficientes para compensar a com-

plexidade, [procurando](#) falhas e também suporte, invocando a [curiosidade](#) — uma vez que aceitam alguma premissa central. Sem a [falácia da conjunção](#), ainda pode haver um efeito halo, mas não haverá uma espiral da autodestruição feliz.

Mesmo nas coisas mais fascinantes do universo conhecido, um racionalista perfeito que exige exatamente a evidência necessária para cada afirmação adicional (positiva) não experimentará nenhuma ressonância emocional. Não podemos alcançar esse nível de perfeição, mas podemos nos aproximar o suficiente do racionalismo para evitar que nossa felicidade saia do controle.

Os casos realmente perigosos são aqueles em que qualquer crítica a qualquer afirmação positiva sobre a Grande Ideia parece ruim ou é socialmente inaceitável. [Os argumentos se tornam soldados, qualquer afirmação positiva é um soldado do nosso lado, e questionar esses soldados é considerado traição.](#) Então, a reação em cadeia se torna supercrítica. Falaremos mais sobre isso posteriormente.

Stuart Armstrong oferece [conselhos](#) intimamente relacionados:

Divida sua Grande Ideia em ideias menores independentes e trate-as como tais.

Por exemplo, um marxista dividiria a Grande Ideia de Marx em uma teoria do valor do trabalho, uma teoria das relações políticas entre as classes, uma teoria dos salários, uma teoria sobre o estado político final da humanidade. Cada uma delas deve ser avaliada independentemente, e a verdade ou falsidade de uma não deve afetar as outras. Se conseguirmos fazer isso, estaremos protegidos da espiral, pois cada teoria é muito específica para iniciar uma espiral por si só.

Isso é metaforicamente como impedir a junção de massas subcríticas de plutônio. Três Grandes Ideias têm muito menos probabilidade de nos enlouquecer do que uma única Grande Ideia. O conselho de Armstrong também ajuda a promover a especificidade: assim que alguém diz: “Publicar artigos suficientes pode curar o câncer”, você pergunta: “Isso é um benefício do método experimental e, se sim, em qual estágio do processo experimental o câncer é curado? Ou é um benefício da ciência como um processo social e, se sim, depende de cientistas individuais que desejam curar o câncer ou podem ser egoístas?” Esperançosamente, isso o afastará do julgamento positivo ou negativo e o levará a perceber a confusão e a falta de embasamento.

Resumindo, você pode evitar uma espiral da autodestruição feliz ao:

- Dividir a Grande Ideia em partes;
- Tratar cada detalhe adicional como algo exigente;
- Considerar as especificidades da cadeia causal em vez de focar apenas nos sentimentos positivos ou negativos;
- Não se deixar levar pela felicidade proveniente de afirmações que “não podem ser provadas como erradas”;

mas não ao:

- Se recusar a admirar muito qualquer coisa;
- Não realizar uma busca tendenciosa por pontos negativos até se sentir infeliz novamente; ou
- Não tentar forçar uma ideia a se encaixar em uma perspectiva.

## 108 - Supercriticidade acrítica



De vez em quando, vemos discussões sobre se o ateísmo é uma “religião”. Como mencionei em [Propósito e Pragmatismo](#), discutir o significado de uma palavra quase sempre significa perder de vista a questão original. Como esse argumento surge?

Um ateu expressa sua opinião, culpando a “religião” pela Inquisição, Cruzadas e conflitos envolvendo o Islã. O religioso pode responder: “Mas o ateísmo também é uma religião, pois você tem crenças sobre Deus; você acredita que Ele não existe”. O ateu retruca: “Se o ateísmo é uma religião, então não colecionar selos é um passatempo”, e a discussão começa.

Ou o religioso pode dizer: “Mas horrores semelhantes foram infligidos por Stalin, que era ateu e suprimiu igrejas em nome do ateísmo; logo, você está errado em culpar a religião pela violência”. O ateu pode cair na tentação de responder com a [falácia do escocês de verdade](#), dizendo: “A religião de Stalin era o comunismo”. O religioso então responde: “Se o comunismo é uma religião, então o *fandom*<sup>47</sup> de Star Wars é um governo”, e a discussão continua.

Deveríamos definir uma pessoa “religiosa” como alguém com opinião definida sobre a existência de pelo menos um Deus, atribuindo uma probabilidade menor que 10% ou maior que 90% à existência de Zeus, por exemplo? Ou uma pessoa “religiosa” deveria ser definida como alguém com uma visão positiva, digamos, uma probabilidade acima de 90%, sobre a existência de pelo menos um Deus? No primeiro caso, Stalin seria considerado “religioso”; no último, “não religioso”. Mas essa é a forma errada de abordar o problema. O que realmente queremos saber — a questão original do debate — é por que, em certos momentos da história, grandes grupos de pessoas foram massacrados e torturados, supostamente em nome de uma ideia. Redefinir uma palavra não muda os fatos históricos.

O comunismo foi uma catástrofe complexa e pode não ter uma única causa, nenhum elo crítico na cadeia causal. Mas se eu tivesse que sugerir um erro fundamental, seria... Bem, vou deixar que Deus fale por mim:

Se teu irmão, filho de teu pai ou de tua mãe, ou teu filho ou filha, ou a esposa que abraças, ou teu amigo mais íntimo, tentar te seduzir em segredo, dizendo: ‘Vamos adorar outros deuses’, deuses que nem tu, nem teus pais conhecem, deuses dos povos ao teu redor, próximos ou distantes, em qualquer lugar do mundo, não deves consentir **nem o ouvir**, não deves ter piedade dele, não deves poupá-lo nem esconder sua culpa. Ao contrário, **deves matá-lo**; tua mão deve ser a primeira a golpeá-lo, e as mãos do povo te seguirão. Deves apedrejá-lo até a morte, ao tentar te desviar de Javé, teu Deus.

— Deuteronômio 13:7–11, ênfase adicionada

Essa também foi a regra de Stalin para o comunismo e de Hitler para o nazismo: se teu irmão disser que Marx está errado, se teu filho disser que os judeus não planejam dominar o mundo, não discuta, não apresente provas, não faça experimentos ou examine a história; entregue-o à polícia secreta.

---

47 NT.: comunidade de fãs.

Sugeri que uma chave para [resistir](#) a uma [espiral da autodestruição](#) afetiva é o princípio dos detalhes onerosos — questionar os detalhes de cada afirmação positiva sobre a Grande Ideia. (Não é um conselho trivial. As pessoas esquecem disso ao ouvir um futurista detalhando as maravilhas do amanhã, ou mesmo ao considerar sua ideia favorita.) Isso não eliminaria o efeito [halo](#), mas reduziria a ressonância, para que uma afirmação aparentemente boa gere menos de uma afirmação adicional, em média.

O oposto, que leva ao efeito halo supercrítico, é quando parece errado contestar qualquer afirmação positiva sobre a Grande Ideia. [A política é a assassina da mente](#). Argumentos são soldados. Uma vez que você escolheu um lado, deve apoiar todas as afirmações favoráveis e argumentar contra as desfavoráveis. Caso contrário, é como ajudar o inimigo ou trair seus amigos.

Se...

- ... você acha que contradizer alguém com uma afirmação imperfeita a favor da evolução ajuda os criacionistas;
- ... você sente que recebe crédito espiritual por cada coisa boa que diz sobre Deus e discutir isso atrapalha seu relacionamento com Ele;
- ... você sente que os outros não vão gostar de você por “não apoiar nossas tropas” se você argumentar contra a última guerra;
- ... dizer algo contra o comunismo te leva a ser ~~apedrejado até a morte~~ morto;
- ... então a espiral da morte afetiva se tornou supercrítica. Agora é uma Espiral da Autodestruição Superfeliz.

Não é a religião, em si, a categoria-chave para a pergunta: “O que causa o massacre?” [A melhor distinção que eu já ouvi](#) entre visões de mundo “sobrenaturais” e “naturalistas” é que a primeira afirma a existência de substâncias mentais ontologicamente básicas, como espíritos, enquanto a segunda reduz fenômenos mentais a elementos não mentais. Focar nisso como a fonte do problema favorece o excepcionalismo religioso. Afirmações sobrenaturalistas valem a pena distinguir, pois sempre se mostram erradas. Mas ainda é apenas um tipo de erro.

## 109 - Resfriamento evaporativo das crenças de grupos



Os primeiros estudiosos de seitas ficaram surpresos ao descobrir que, quando estes enfrentam grandes reveses — como uma profecia não cumprida ou a revelação de uma falha moral do líder —, muitas vezes se tornam ainda mais fortes, com crenças e fanatismo intensificados. As Testemunhas de Jeová, por exemplo, esperavam o Armagedom em 1975, com base em cálculos bíblicos. O ano veio e se foi, sem qualquer evento. A seita unariana, ainda ativa, sobreviveu à não chegada de uma frota espacial intergaláctica em 27 de setembro de 1975.

Por que uma crença coletiva se fortaleceria diante de evidências contrárias tão contundentes?

A interpretação convencional se baseia na dissonância cognitiva. Quando pessoas tomam ações “irreversíveis” em prol de uma crença — como doar tudo esperando a chegada de discos voadores —, não conseguem admitir estarem erradas. O desafio à crença gera uma enorme dissonância; elas precisam encontrar pensamentos que a reforcem para lidar com o choque, tornando-se mais fanáticas. Nessa interpretação, o aumento do fanatismo do grupo resulta do aumento do fanatismo individual.

Observando uma *applet* Java que simula o resfriamento evaporativo para formar um condensado de Bose-Einstein, pensei que outra força poderia aumentar o fanatismo. O resfriamento evaporativo cria uma barreira de energia potencial em torno de átomos quentes. A energia térmica é estatística — nem todos os átomos se movem na mesma velocidade. A energia cinética de cada átomo varia conforme colidem. Se a barreira de energia potencial for um pouco maior que a energia térmica média, ocasionalmente, um átomo terá energia suficiente para escapar. Ao escapar, leva consigo uma grande quantidade de energia cinética, diminuindo a energia média do grupo. Assim, o grupo fica mais frio que a barreira ao seu redor. Brincar com a *applet* pode ajudar a visualizar isso.

No clássico *When Prophecy Fails* (Quando a profecia falha), de Festinger, Riecken e Schachter, um membro da seita saiu após a não chegada da nave espacial [\[1\]](#). Quem sai primeiro? Um membro comum? Ou um mais cético, que antes moderava os mais fanáticos?

Após a saída dos membros com maior energia cinética, as discussões restantes serão entre fanáticos extremos de um lado e fanáticos um pouco menos extremos do outro, com o consenso no “meio”.

E qual seria a analogia com o colapso para formar um condensado de Bose-Einstein? Bem, não precisamos ir tão longe. Mas lembre-se da analogia da reação em cadeia para a espiral da morte afetiva: quando um grupo expulsa as vozes moderadas, as pessoas que se incentivam e suprimem discordâncias podem aumentar o fanatismo médio. (Sem analogia termodinâmica aqui, a menos que alguém invente uma arma nuclear que explode ao esfriar.)

Quando o caso de Ayn Rand com Nathaniel Branden veio à tona, muitos objetivistas seguiram Branden, defendendo um “sistema aberto” de objetivismo menos ligado a Rand. Quem ficou com ela? Os que realmente acreditavam — e talvez alguns indecisos que, sem as vozes moderadas, só ouviram um lado. Isso pode explicar como o Instituto Ayn Rand é (supostamente) mais fanático após a separação do que o grupo original sob Branden e Rand.

Anos atrás, participei de uma lista de discussão transhumanista onde um pequeno grupo de “transumanistas social-democratas” insultava ferozmente os libertários. A maioria saiu e outros pararam de participar. O grupo restante se inclinou mais à esquerda. Foi intencional? Provavelmente não, pois não creio que

soubessem muito de psicologia. (Aliás, não me lembro de ter visto a analogia do resfriamento evaporativo em outro lugar, mas isso não significa que não tenha sido mencionada antes.) No máximo, podem ter pensado em se tornar “peixes maiores em um lago menor”.

Por isso, é importante tolerar a dissidência. Espere bem mais do que parece justificado antes de expulsar alguém. Se você se livrar dos discordantes, a posição do grupo muda e outro se torna o esquisito. Se o expulsar também, você caminha para se tornar um condensado de Bose-Einstein e, bem, explodir.

Por outro lado, Thomas Kuhn acreditava que a ciência precisa se tornar um “paradigma”, com linguagem técnica que exclui os de fora, para funcionar. No início, segundo Kuhn, os adeptos se esforçam para tornar seu trabalho compreensível. Mas a ciência só progride como disciplina técnica quando abandona a exigência de acessibilidade externa, e os cientistas assumem a familiaridade com um vasto núcleo de material técnico em suas comunicações. Isso soa cínico em relação à compreensão pública da ciência, mas há verdade nisso.

Minha teoria sobre moderação na internet é que você deve excluir *trolls* e spam para iniciar uma conversa. Para trabalhar, exclua pessoas gentis, mas desinformadas, de listas técnicas. Uma conversa aberta na internet degenera rapidamente. Cuidado ao excluir *trolls* articulados, pois eles legitimam discordâncias menos extremas. Mas não tenha tantos que eles comecem a discutir entre si ou dominar as conversas. Se há alguém famoso por discordar de tudo, quem discorda moderadamente não parecerá o único a se destacar. Essa teoria pode não ter me servido bem, então, tome-a com cautela.

## Referências

[1] Leon Festinger, Henry W. Riecken, and Stanley Schachter, *When Prophecy Fails: A Social and Psychological Study of a Modern Group That Predicted the Destruction of the World* (Harper-Torchbooks, 1956).

## 110 - Quando ninguém ousa pedir moderação



Certa manhã, levantei, liguei o computador e meu cliente de e-mail Netscape baixou as notícias do dia. Naquele dia, a notícia era que dois aviões sequestrados haviam atingido o World Trade Center.

Meus três primeiros pensamentos foram:

*“Acho que realmente estou vivendo no futuro.”*

*“Ainda bem que não foi um ataque nuclear.”*

E então...

*“A reação exagerada a isso será dez vezes pior que o evento em si.”*

O termo “dez vezes pior” foi um eufemismo. Nem eu imaginei o quão ruim as coisas ficariam. Esse é o desafio do pessimismo: é difícil ter expectativas baixas o suficiente para ser agradavelmente surpreendido com a mesma frequência que se é desagradavelmente surpreendido.

No entanto, percebi que todos enfatizariam o quão terrível foi o evento, e ninguém ousaria ser a voz da moderação, da resposta proporcional. Inicialmente, em 11 de setembro, estimava-se 6 mil mortos. Qualquer político que dissesse “6 mil mortes são apenas 1/8 das mortes anuais por acidentes de carro nos EUA” seria convidado a renunciar.

Não, o 11 de setembro não foi um dia bom. Mas se todos ganham pontos por enfatizar a dor, e ninguém pede moderação na retaliação, a reação será desproporcional.

Este é o reflexo ainda mais sombrio da espiral do ódio — [a espiral da autodestruição feliz](#). Quem ataca o *Inimigo* é patriota; quem questiona uma única declaração negativa sobre o *Inimigo* é traidor. Mas, assim como a maioria das declarações complexas é falsa, a maioria das coisas negativas sobre alguém, mesmo a pior pessoa do mundo, também é falsa.

A melhor ilustração foi a afirmação de que [os sequestradores suicidas eram covardes](#). Um pouco de bom senso, por favor! É preciso coragem para voar um avião contra um prédio. Covardia não estava entre seus pecados. Mas parece que qualquer coisa negativa sobre um terrorista, por mais absurda, deve ser verdade. Eu ganharia mais pontos se acusasse a Al-Qaeda de matar John F. Kennedy? Ou de serem stalinistas? Sério, covardes?

Sim, é importante entender que os sequestradores não eram covardes, e não apenas para entender a psicologia do inimigo. Há um grande perigo em perpetuar espirais de ódio. É arriscado permitir que existam alvos, sejam judeus ou Hitler, sobre os quais é mais aceitável dizer coisas negativas do que precisas.

Quando a força de defesa tem milhares de aeronaves e soldados, devemos considerar que o próprio sistema imunológico pode causar mais dano do que 19 homens e quatro aviões. Os EUA gastaram bilhões e perderam milhares de vidas, atirando no próprio pé com mais eficiência do que qualquer terrorista.

Se os EUA tivessem ignorado o 11 de setembro — simplesmente reconstruindo o prédio —, teria sido melhor. Mas não era uma opção política viável. Mesmo que alguém percebesse que a resposta seria pior que a doença, os políticos não tinham escolha a não ser cair na armadilha da Al-Qaeda para preservar suas carreiras. Quem defende uma resposta mais intensa é patriota. Quem questiona as afirmações patrióticas, é traidor.

Inicialmente, houve respostas mais ponderadas ao 11 de setembro do que eu imaginava. Vi um congressista — não lembro quem — dizer: “Esquecemos que o objetivo principal do governo não é a economia, nem a saúde, mas defender o país”. Fiquei surpreso por ser raro um político dizer algo que não seja apenas busca por aplausos. O impacto emocional deve ter sido grande para um congressista fazer uma afirmação tão realista.

Mas, em dois dias, o impacto genuíno se dissipou e a preocupação com a imagem retomou o controle. A espiral de escalada tomou conta. Quando a moderação se torna inaceitável, não importa onde o discurso comece, a raiva e a loucura só aumentam.



## 111 - O experimento da caverna dos ladrões



Alguma vez, quando criança, você se perguntou se seu acampamento de verão tinha um propósito secreto, como um experimento científico com os monitores observando seu comportamento?

Eu também não.

Mas se tivéssemos lido o estudo [Intergroup Conflict and Cooperation: The Robbers Cave Experiment](#) (Conflito e Cooperação Intergrupar: O Experimento da Caverna dos Ladrões) de Sherif, Harvey, White, Hood e Sherif [1], talvez fôssemos mais paranoicos. Esse estudo investigou as causas e soluções para o conflito entre grupos. Os participantes, ou “campistas”, eram 22 meninos de 11 e 12 anos, de 22 escolas diferentes em Oklahoma City, de famílias protestantes estáveis de classe média, com bom desempenho escolar e QI médio de 112. Os pesquisadores buscaram grupos de meninos o mais semelhantes e adaptados possível.

O experimento, após a Segunda Guerra Mundial, visava investigar as causas — e possíveis soluções — do conflito intergrupar. Como eles iniciariam um conflito? Bem, dividiram os 22 meninos em dois grupos de 11, e isso foi o suficiente.

O plano original tinha três etapas. Na Etapa 1, cada grupo se instalaria sem saber da existência do outro. No final da Etapa 1, seriam gradualmente informados sobre o outro grupo. Na Etapa 2, competições com prêmios colocariam os grupos em conflito.

Mas a Etapa 2 foi desnecessária. A hostilidade surgiu assim que cada grupo soube do outro: eles estavam usando nosso acampamento, nosso campo de beisebol. No primeiro encontro, já trocaram insultos. Se autodenominaram *Rattlers* e *Eagles*<sup>48</sup> (nomes desnecessários quando eram o único grupo).

Quando os concursos foram anunciados, a rivalidade se intensificou. O espírito esportivo durou dois dias, mas logo se deteriorou.

Os *Eagles* roubaram e queimaram a bandeira dos *Rattlers*. Os *Rattlers* invadiram a cabana dos *Eagles* e roubaram a calça jeans do líder, que pintaram de laranja e usaram como bandeira, com a frase “O último dos *Eagles*”. Os *Eagles* revidaram, invadindo a cabana dos *Rattlers*, virando camas e espalhando terra. Voltaram para sua cabana, se fortificaram e prepararam armas (meias com pedras). Após os *Eagles* vencerem a última competição da Etapa 2, os *Rattlers* invadiram sua cabana e roubaram os prêmios. Isso virou uma briga que os pesquisadores tiveram que interromper. Os *Eagles* contaram a história como uma grande vitória — eles perseguiram os *Rattlers* “até a metade do caminho de volta” (o que não aconteceu).

Cada grupo desenvolveu um estereótipo negativo do outro e um positivo de si mesmo. Os *Rattlers* xingavam muito. Os *Eagles*, após uma vitória, concluíram que venceram por suas orações e que os *Rattlers* perderam pelos palavrões. Decidiram parar de xingar e evitar os *Rattlers*. Os *Eagles* se viam como corretos e morais, enquanto os *Rattlers* se viam como durões.

Os membros de um grupo tapavam o nariz quando o outro grupo passava.

Na Etapa 3, os pesquisadores tentaram reduzir o atrito. O mero contato não funcionou. Assistir a eventos juntos, como fogos de artifício no 4 de julho, resultou em guerra de comida.

---

48 NT.: Cascavéis (*Rattlers*) e águias (*Eagles*).

Quer adivinhar o que funcionou?

(Alerta de spoiler...)

Disseram aos meninos que poderia haver falta de água no acampamento devido a um problema no sistema — talvez causado por vândalos. (O Inimigo Externo, um truque clássico.)

Organizaram uma busca entre o acampamento e o reservatório, com quatro equipes (inicialmente, com membros de ambos os grupos). Se nada fosse encontrado, todos se encontrariam na caixa d'água. Nada foi encontrado, e os grupos viram que não havia água nas torneiras. Discutiram o problema, bateram na caixa d'água, encontraram uma escada, verificaram se estava cheia e acharam um saco obstruindo a torneira. Todos a limparam. Sugestões de ambos os grupos foram dadas e implementadas.

Quando a torneira foi limpa, os *Rattlers*, que tinham cantis, deixaram os *Eagles* beberem primeiro (eles não tinham cantis). Sem insultos, nem mesmo o típico “damas primeiro”.

A rivalidade não acabou. Houve outra guerra de comida na manhã seguinte. Mas algumas tarefas que exigiam cooperação entre os dois grupos — como, por exemplo, ligar um caminhão enguiçado — foram eficazes. No final, os *Rattlers* usaram US\$5 ganhos em um concurso para comprar malte para todos.

O experimento da Caverna dos Ladrões ilustra a psicologia dos bandos de caçadores-coletores, ecoando ao longo do tempo, tão perfeitamente quanto qualquer experimento já concebido pela ciência social.

Qualquer semelhança com a política moderna é pura imaginação. (Às vezes, eu penso que a segunda maior necessidade da humanidade é um supervilão.

## Referências

[1] Muzafer Sherif et al., “Study of Positive and Negative Intergroup Attitudes Between Experimentally Produced Groups: Robbers Cave Study,” Unpublished manuscript (1954).

## 112 - Toda Causa Quer Ser Uma Seita



Cade Metz, do *The Register*, [alegou recentemente](#) que uma lista secreta de e-mail dos principais administradores da Wikipédia se tornou obcecada em banir todos os críticos e possíveis críticos da plataforma. Isso incluiu o banimento de um usuário produtivo quando um administrador — apenas devido à produtividade — se convenceu de que o usuário era um espião enviado pela Wikipedia Review. Metz também afirmou que as figuras importantes da Wikipédia se uniram para se defender. (Ainda não investiguei essas alegações. Um aceno para Eugen Leitl.)

Será que existe uma falha moral profunda em sistematizar o conhecimento do mundo, que levaria os defensores dessa causa à obsessão? Talvez apenas pessoas com tendências totalitárias inatas tentariam se tornar a autoridade mundial em tudo—

Alerta de [viés de correspondência](#)! (Viés de correspondência: inferir características individuais únicas com base em comportamentos que podem ser totalmente explicados pela situação. Quando vemos alguém chutar uma máquina de venda automática, pensamos que a pessoa está “com raiva”, mas quando chutamos a máquina, é porque o ônibus atrasou, o trem se adiantou e a máquina engoliu nosso dinheiro.) Se as alegações sobre a Wikipédia forem verdadeiras, elas se explicam pela natureza humana comum, não por uma natureza humana extraordinária.

A [dicotomia entre grupo interno e externo](#) é parte da natureza humana comum, assim como as [espirais da autodestruição feliz](#) e as [espirais de ódio](#). Uma causa nobre não precisa ter uma falha profunda oculta para seus seguidores se tornarem um grupo sectário. Basta que os seguidores sejam humanos. O resto ocorre naturalmente, como comida que estraga na geladeira sem energia.

Assim como todo gradiente de temperatura busca se igualar e todo software quer se tornar uma coleção de remendos, toda Causa quer ser uma seita. É um estado de alta entropia para o qual o sistema tende, um atrator na psicologia humana. Isso pode não ter relação com a nobreza da Causa. Podemos pensar que uma Causa boa compartilharia sua bondade em todos os aspectos das pessoas associadas a ela — que os seguidores seriam menos suscetíveis a jogos de status, vieses de grupo, espirais afetivas e líderes como deuses. Mas acreditar em uma ideia verdadeira não elimina o [efeito halo](#). Uma Causa nobre não transformará seus seguidores em algo além de humanos. Há muitas ideias ruins que podem causar danos, mas não é necessariamente isso que está acontecendo.

Todo grupo com um objetivo incomum — bom, mau ou bobo — tenderá a se tornar uma seita, a menos que se esforce para resistir. Você pode manter sua casa mais fria que o exterior, mas precisa do ar condicionado ligado, e assim que a energia acabar — ou seja, desistir da luta contra a entropia —, as coisas voltam ao “normal”.

Houve um grupo que se tornou meio sectário cujo lema era “Racionalidade! Razão! Realidade objetiva!” (Mais sobre isso depois.) Rotular a Grande Ideia de “racionalidade” não te protegerá mais do que colocar uma placa “Frio!” em sua casa. Você ainda precisa ligar o ar condicionado — gastar energia para reverter a tendência natural ao sectarismo. Adorar a racionalidade não te tornará são, assim como adorar a gravidade não te permite voar. Você não pode falar com a termodinâmica nem rezar para a teoria da probabilidade. Você pode usá-las, mas não se juntar a elas como um grupo.

O sectarismo é quantitativo, não qualitativo. A pergunta não é “É uma seita ou não?”, mas “Quanto sectarismo e onde?”. Mesmo na ciência, o arquétipo de uma Causa Verdadeiramente Nobre, vemos as

fronteiras da luta contra a entropia da seita, onde a linha de batalha oscila. Revistas científicas aceitam mais artigos de autores conhecidos ou de instituições renomadas, em comparação com autores desconhecidos? Quanta crença vem da autoridade e quanta da experimentação? Quais revistas usam revisão por pares cega e qual sua eficácia?

Cito este exemplo, em vez das vagas acusações [costumeiras](#) de que “cientistas não estão abertos a novas ideias”, ao ilustrar uma linha de batalha — onde a psicologia humana é repelida, onde a entropia da seita é dissipada. Claro, isso gera calor residual.

Este ensaio não é um catálogo de técnicas contra o sectarismo. [Algumas](#) já foram mencionadas, [outras](#) virão depois. Aqui, destaco que o valor de uma Causa não significa menos esforço na resistência ao atrator da seita. E apontar as linhas de batalha não significa admitir que sua Nobre Causa seja indigna. Talvez você pense que, se a pergunta fosse “É uma seita?”, você teria que responder “Não” para não trair sua Causa. Mas isso é como dividir motores em “perfeitamente eficientes” e “ineficientes”, em vez de medir o desperdício.

Por outro lado, se você acredita que foi a impureza inerente de outras Causas que as fez falhar, se você ri da insanidade das “vítimas de seitas”, se você acha que seitas são lideradas e habitadas por mutantes, então você não se esforçará para combater a entropia — para resistir ser humano.

## 113 - Guardiões da verdade



A crítica, por vezes, é direcionada aos racionalistas: “A Inquisição achava que tinha a verdade! Claramente, essa coisa de ‘verdade’ é perigosa.” Existem várias respostas óbvias, como “Se você acha que ter a verdade te autoriza a torturar e matar, você está cometendo um erro que não tem nada a ver com epistemologia.” Ou “Então, aquela declaração histórica que você acabou de fazer sobre a Inquisição, é verdadeira?”

A estupidez ao contrário não é inteligência: “Se o seu computador atual parar de funcionar, você não pode concluir que tudo no sistema atual está errado e que você precisa de um novo sistema sem um processador AMD, uma placa de vídeo ATI... mesmo que seu sistema atual tenha todas essas coisas e não funcione. Talvez você só precise de um novo cabo de alimentação.” Chegar a uma conclusão ruim requer apenas um passo errado, não todos os passos errados. Os inquisidores acreditavam que  $2 + 2 = 4$ , mas isso não era a fonte de sua loucura. Talvez o realismo epistemológico também não fosse o problema?

Parece plausível que, se a Inquisição fosse composta de relativistas, professando que nada era verdade e nada importava, eles teriam mostrado menos entusiasmo por sua tortura. Eles também teriam demonstrado menos entusiasmo se tivessem sido lobotomizados. Acredito que seja uma analogia válida. E ainda assim... acredito que a atitude da Inquisição em relação à verdade desempenhou um papel. A Inquisição acreditava que existia algo chamado verdade e que isso era importante; assim como Richard Feynman. Mas os inquisidores não eram buscadores da verdade. Eles eram os Guardiões da Verdade.

Certa vez, li um argumento (não consigo encontrar a fonte) de que um componente crucial de um zeitgeist é se ele localiza seus ideais no futuro ou no passado. Quase todas as culturas anteriores ao Iluminismo acreditavam em uma Perda de Prestígio — que as coisas já foram perfeitas em um passado distante, então ocorreu uma catástrofe e tudo desmoronou lentamente desde então:

Na época em que a vida na Terra era plena... Eles se amavam e não sabiam que aquilo era “amor ao próximo”. Eles não enganavam ninguém, mas não sabiam serem “pessoas de confiança”. Eles eram confiáveis e não sabiam que aquilo era “boa-fé”. Viviam juntos, dando e recebendo livremente, e não sabiam serem generosos. Por isso, seus atos não foram registrados. Eles não fizeram história<sup>49</sup>.”

—*The Way of Chuang Tzu* (O caminho de Chuang Tzu), traduzido para o inglês por [Thomas Merton](#) [1].

A idade perfeita do passado, de acordo com nossa melhor evidência antropológica, nunca existiu. No entanto, uma cultura que enxerga a vida fluindo inexoravelmente para baixo é muito diferente de uma cultura em que você pode alcançar alturas sem precedentes. (Emprego o termo “cultura” e não “sociedade” porque é possível ter várias subculturas numa sociedade.)

Poderíamos dizer que a diferença entre a Inquisição e Richard Feynman era que a Inquisição acreditava que possuía a verdade, enquanto Richard Feynman buscava a verdade. No entanto, essa afirmação não é totalmente defensável, pois, sem dúvida, havia algumas verdades que Richard Feynman acreditava também possuir, como “o céu é azul” ou “ $2 + 2 = 4$ ”.

---

49 NT.: original em chinês, traduzido para o inglês por Thomas Merton. *In the age when life on Earth was full... They loved each other and did not know that this was “love of neighbor.” They deceived no one yet they did not know that they were “men to be trusted.” They were reliable and did not know that this was “good faith.” They lived freely together giving and taking, and did not know that they were generous. For this reason their deeds have not been narrated. They made no history.*

Sim, existem certas verdades efetivas na ciência. A Teoria da Relatividade Geral pode ser refutada por futuras descobertas da física — embora de forma alguma isso preveja que o Sol orbitará Júpiter. A nova teoria deve incorporar as previsões bem-sucedidas da antiga teoria, não as contradizer.

Embora a teoria da evolução opere em um nível de organização superior ao dos átomos, nenhum conhecimento obtido sobre quarks anula os princípios do darwinismo, da teoria celular da biologia, da teoria atômica da química ou de várias outras descobertas brilhantes cuja validade está consolidada e indiscutível.

São essas “verdades absolutas”? Não no sentido de possuir uma probabilidade literalmente igual a 1,0. No entanto, são casos em que a ciência basicamente considera que possui a verdade. No entanto, os cientistas não torturam pessoas que questionam a teoria atômica da química. Por quê? Será porque não acreditam que a certeza justifica a tortura? Bem, essa é a diferença superficial, mas por que os cientistas não acreditam nisso?

Será porque a química não impõe uma penalidade sobrenatural de tortura eterna por não acreditar na teoria atômica da química? No entanto, voltamos e nos perguntamos: “Por quê?” Por que os químicos não acreditam que você irá para o inferno se não acreditar na teoria atômica?

Será porque as revistas científicas não publicarão seu artigo até você obter uma observação experimental sólida do inferno? Ainda assim, muitos cientistas conseguem suprimir seu reflexo cético na hora que querem. Por que os químicos não possuem uma seita que argumenta que os não químicos vão para o inferno, já que muitos deles são cristãos de qualquer maneira?

Perguntas como essa não têm respostas simples devido a um único fator. No entanto, eu diria que um dos elementos está relacionado a adotar uma postura produtiva em relação à verdade, em oposição a uma postura defensiva em relação a ela.

Quando você se torna o Guardião da Verdade, sua única contribuição útil para a Verdade é protegê-la. Quando você está buscando ganhar o Prêmio Nobel de Química descobrindo o próximo benzeno ou fulereno, alguém que desafia a teoria atômica não representa tanto uma ameaça à sua visão de mundo quanto uma perda de tempo.

Quando você é um Guardião da Verdade, tudo o que você pode fazer é tentar evitar o inevitável declínio em direção à entropia, eliminando qualquer coisa que se desvie da Verdade. Se houver uma maneira de bombear contra a entropia, gerando novas crenças verdadeiras juntamente com um pouco de calor residual, essa mesma bomba pode manter a verdade preciosa viva sem a necessidade de uma polícia secreta. Na química, você pode replicar experimentos e ver por si mesmo — e isso mantém a valiosa verdade viva sem a necessidade de violência.

E não é tão terrível se cometermos um erro em algum lugar — acabamos acreditando em uma pequena inverdade por um tempo — porque amanhã podemos recuperar o terreno perdido.

Mas todo esse truque só funciona porque o método experimental é um “critério de validade” e não um mero “critério de comparação”. Assim como os experimentos podem recuperar a verdade sem a necessidade de autoridade, eles também podem anular a autoridade e criar novas crenças verdadeiras onde antes não existiam.

Onde existem critérios de validade que não são apenas critérios de comparação, podem ocorrer mudanças que são melhorias, ao invés de ameaças. Onde há apenas critérios de comparação, onde não há como contornar a autoridade, também não há como resolver um desacordo entre as autoridades. A menos que seja pela extinção. As armas mais poderosas vencem.

Não pretendo fornecer uma visão abrangente e completa da história. Pretendo, sim, destacar uma diferença psicológica fundamental entre ver sua grande causa na vida como proteção, guarda e preservação, em oposição a descobrir, criar e melhorar. A direção “para cima” do tempo aponta para o passado ou para o futuro? Essa distinção lança sua sombra sobre tudo, estendendo seus tentáculos por todos os lugares.

É por isso que sempre insisto, por exemplo, que, se você pretende discutir a “ética da IA”, é melhor abordar como melhorar a situação atual usando IA, em vez de apenas evitar que várias coisas deem errado. Ao adotar critérios meramente comparativos, você começa a perder o controle de seus ideais — perde de

vista o que é certo e errado e começa a ver apenas o que é “diferente” e “igual”.

Também argumentaria que essa diferença psicológica fundamental é uma das razões pelas quais um campo acadêmico que para de fazer progresso ativo tende a se tornar mesquinho (pelo menos pelos padrões refinados da ciência). A caça às bruxas é moderada pelos padrões históricos; a maioria dos sistemas de crença com postura defensiva já foi levada ao extremo. Quando a quantidade de novas descobertas diminui, não há mais nada para se fazer durante o dia, exceto caçar os hereges.

Para obter os maiores benefícios para a saúde mental da postura de descobrir/criar/melhorar, é preciso realmente progredir, e não apenas esperar por isso.

## **Referências**

[1] Zhuangzi and Thomas Merton, *The Way of Chuang Tzu* (New Directions Publishing, 1965).

## 114 - Os guardiões do pool genético



Como qualquer cidadão educado do século XXI, você provavelmente já ouviu falar da Segunda Guerra Mundial. Talvez se lembre de que Hitler e os nazistas planejavam levar adiante um processo romantizado de evolução, para criar uma nova raça superior, super-homens, mais fortes e inteligentes do que qualquer coisa que já existiu.

Na verdade, este é um equívoco comum. Hitler acreditava que o super-homem ariano já havia existido — o estereótipo nórdico, a fera loira de olhos azuis —, mas fora corrompido pela mistura com raças impuras. Houve uma Perda de Prestígio racial.

É revelador o quanto o conceito de progresso permeia a civilização ocidental, ao ponto de, ao se falar sobre a eugenia nazista, ouvirmos: “Eles tentaram criar um super-humano”. Você, caro leitor — se fracassasse tanto a ponto de apoiar a eugenia coercitiva, tentaria criar um super-humano. Porque você coloca seus ideais no futuro, não no passado. Porque você é criativo. A ideia de reproduzir um arquétipo nórdico de mil anos atrás nem sequer lhe ocorreria como uma possibilidade — o quê, só os vikings? É só isso? Se você fracassasse tanto a ponto de matar, com certeza tentaria alcançar patamares nunca atingidos, ou seria um desperdício, não é mesmo? Bem, essa é uma das razões pelas quais você não é um nazista, caro leitor.

É revelador o quão difícil é para pessoas relativamente saudáveis se colocarem no lugar de pessoas doentes, quando nos falamos dos nazistas e distorcem a história para retratá-los como transumanistas defeituosos.

Foram os comunistas os transumanistas defeituosos. “Novo Homem Soviético” e tudo mais. Os nazistas eram definitivamente os bioconservadores da história.



## 115 - Os guardiões de Ayn Rand



Para os céticos, a ideia de que a razão possa levar a uma seita é absurda. As características de uma seita estão totalmente em desacordo com a razão. No entanto, como vou demonstrar, não apenas isso pode acontecer, mas já aconteceu, e envolve um grupo que seria considerado a seita mais improvável da história. É uma lição sobre o que acontece quando a busca pela verdade se torna mais importante do que a própria verdade...<sup>50</sup>

—Michael Shermer, [The Unlikeliest Cult in History](#) (A seita mais improvável da história) [1]

Michael Shermer, em seu artigo *The Unlikeliest Cult in History* (A Seita Mais Improvável da História), está, em minha opinião, fornecendo uma explicação abrangente do Objetivismo. Permitam-me expandir essa ideia.

Os romances de Ayn Rand glorificam a tecnologia, o capitalismo, o desafio individual ao sistema, o governo limitado, a propriedade privada e o egoísmo. Seu último herói fictício, John Galt, é um cientista que inventou uma nova forma de energia renovável e barata, mas se recusa a compartilhá-la com o mundo, pois os lucros seriam apenas roubados para sustentar governos corruptos.

E então, de alguma forma, tudo se transformou em um “sistema fechado” moral e filosófico, com Ayn Rand no centro. Vale ressaltar que o termo “sistema fechado” não é uma acusação minha, mas sim como o Instituto Ayn Rand descreve o Objetivismo.

O Objetivismo é definido pelas obras de Ayn Rand. Agora que Rand está morta, o Objetivismo é considerado um sistema fechado. Se você discordar de qualquer aspecto das obras de Rand, não poderá ser considerado um objetivista.

Max Gluckman certa vez disse: “Uma ciência é qualquer disciplina na qual o tolo desta geração pode ir além do ponto alcançado pelo gênio da geração anterior”. A ciência avança ao derrubar seus heróis, como quando Newton foi superado por Einstein. Todo jovem físico sonha em se tornar o novo campeão, aquele que os futuros físicos sonharão em desafiar.

O ídolo filosófico de Ayn Rand era Aristóteles. Embora Aristóteles possa ter sido um talento jovem em matemática há 2.350 anos, a matemática progrediu significativamente desde então. A teoria da probabilidade bayesiana é a lógica quantitativa da qual a lógica qualitativa de Aristóteles é um caso especial. No entanto, não há evidências de que Ayn Rand tenha conhecimento da teoria da probabilidade bayesiana ao escrever sua obra-prima, “A Revolta de Atlas”. Rand discutiu sobre “racionalidade”, mas falhou em se familiarizar com pesquisas modernas sobre heurísticas e vieses. Como alguém pode se autodenominar um mestre racionalista e, ao mesmo tempo, desconhecer esses tópicos fundamentais?

“Espere um minuto”, contesta o leitor, “isso não é justo! “A Revolta de Atlas” foi publicado em 1957! Naquela época, quase ninguém sabia sobre Bayes.” Ora. Em seguida, você vai me dizer que Ayn Rand morreu em 1982 e não teve a chance de ler *Judgment Under Uncertainty: Heuristics and Biases* (Julgamento Sob Incerteza: Heurísticas e Vieses), publicado no mesmo ano.

---

50 NT.: Texto original em inglês. *For skeptics, the idea that reason can lead to a cult is absurd. The characteristics of a cult are 180 degrees out of phase with reason. But as I will demonstrate, not only can it happen, it has happened, and to a group that would have to be considered the unlikeliest cult in history. It is a lesson in what happens when the truth becomes more important than the search for truth...*

A ciência não é justa. E esse é o ponto. Um aspirante a racionalista em 2007 tem uma enorme vantagem sobre um aspirante a racionalista em 1957. É assim que sabemos que houve progresso.

Para mim, a ideia de abraçar voluntariamente um sistema explicitamente vinculado às crenças de um ser humano já falecido está em algum lugar entre o tolo e o suicida. Afinal, um computador não se torna obsoleto em apenas cinco anos.

A energia que Rand admirava na ciência, no comércio, em cada ferrovia que substituíra uma rota de charrete, em cada arranha-céu construído com uma nova arquitetura - tudo isso advém do princípio de superar os mestres antigos. Como poderia haver ciência se o cientista mais experiente que já existiu já tivesse vivido? Quem elevaria o horizonte de Nova York que Rand tanto admirava se o edifício mais alto já tivesse sido construído?

No entanto, Ayn Rand nunca reconheceu nenhuma superioridade, nem no passado, nem no futuro. Rand, que inicialmente admirava a razão e a individualidade, acabou condenando ao ostracismo qualquer um que ousasse contradizê-la. Shermer:

[Barbara] Branden lembrou uma noite em que um amigo de Rand comentou que gostava da música de Richard Strauss. “Quando ele saiu no final da noite, Ayn disse, numa reação cada vez mais típica: “Agora entendo por que ele e eu nunca podemos ser almas gêmeas de verdade. Nossas visões de vida estão muito distantes.” Muitas vezes, ela não esperava até que o amigo se fosse para fazer esses comentários.<sup>51</sup>

Ayn Rand mudou ao longo do tempo; isso é suspeito.

Rand cresceu na Rússia e testemunhou a Revolução Bolchevique. Aos 21 anos, ela obteve um visto para visitar parentes nos Estados Unidos e nunca mais voltou. É fácil odiar o autoritarismo quando se é vítima dele. É fácil defender a liberdade do indivíduo quando se é oprimido. É preciso uma força muito maior para temer a autoridade ao ter o poder. Quando as pessoas olham para você em busca de respostas, é mais difícil dizer “O que diabos eu sei sobre música? Sou escritor, não compositor” ou “É difícil ver como gostar de uma peça musical pode ser falso”.

Quando é você quem esmaga aqueles que ousam ofendê-lo, o exercício do poder parece muito mais justificável do que quando é você quem está sendo esmagado. Todos os tipos de justificativas excelentes vêm à mente.

Michael Shermer explora em detalhes como ele acredita que a filosofia de Rand acabou se tornando uma seita. Em particular, Shermer sugere (aparentemente) que o Objetivismo falhou porque Rand acreditava que a certeza era possível, enquanto a ciência nunca é certa. Não posso concordar com Shermer nisso. A teoria atômica da química é bastante precisa. No entanto, os químicos não se tornaram uma seita.

Na verdade, acredito que Shermer está sujeito ao [viés de correspondência](#) ao supor que existe uma correlação específica entre a filosofia de Rand e a maneira como seus seguidores formaram uma seita. [Toda causa deseja se tornar uma seita.](#)

Ayn Rand fugiu da União Soviética, escreveu um livro sobre o individualismo que encontrou muitos admiradores e recebeu muitos elogios. Com o tempo, formou um círculo de seguidores entusiastas. À medida que seus admiradores encontravam mais coisas positivas para dizer sobre ela (uma [espiral da autodestruição feliz](#)), ela gostou disso e não os impediu de continuar. Ela se viu no poder de silenciar aqueles que desaprovava e não resistiu à tentação.

Ayn Rand e Nathaniel Branden tiveram um caso extraconjugal em segredo. (Com a permissão de ambos os cônjuges, o que é relevante em minha opinião. Se você quiser transformar isso em um “problema”, é necessário especificar que os cônjuges estavam infelizes - e mesmo assim, não é assunto para estranhos.)

---

51 NT.: Texto original em inglês. [Barbara] Branden recalled an evening when a friend of Rand's remarked that he enjoyed the music of Richard Strauss. “When he left at the end of the evening, Ayn said, in a reaction becoming increasingly typical, ‘Now I understand why he and I can never be real soulmates. The distance in our sense of life is too great.’” Often she did not wait until a friend had left to make such remarks.

Quando a traição de Branden com outra mulher foi revelada, Rand ficou furiosa e o excomungou. Muitos objetivistas se separaram quando o caso se tornou público.

Quem ficou com Rand, em vez de seguir Branden ou abandonar completamente o Objetivismo? Seus apoiadores mais fervorosos. Quem partiu? As vozes anteriores de moderação. (Esse é o [esfriamento evaporativo das crenças de grupo](#).) Desde então, o domínio de Rand sobre o restante de seu círculo foi absoluto, e nenhum questionamento era permitido.

A única coisa extraordinária em todo esse processo é o quão comum é.

Você pode pensar que um sistema de crenças que exalta a “razão”, a “racionalidade” e o “individualismo” teria alguma espécie de imunidade especial, de alguma forma...?

Bem, não foi o que aconteceu.

Funcionou como colocar um adesivo escrito “Frio” em uma geladeira que não estava ligada.

Não havia um esforço ativo necessário para resistir ao declínio em direção à entropia, e a decadência inevitavelmente se seguiu.

E se você chama isso de “a seita mais improvável da história”, está apenas nomeando a realidade de forma desagradável.

Que isso sirva de lição para todos nós: elogiar a “racionalidade” por si só não significa nada. Até mesmo dizer “Você deve justificar suas crenças pela razão, não concordando cegamente com o Grande Líder” acaba executando um pequeno programa automático que simplesmente aceita tudo o que o Grande Líder diz e gera uma justificativa que os demais seguidores consideram razoável.

Então, onde está a verdadeira arte da racionalidade? Está em estudar a matemática da teoria da probabilidade e da teoria da decisão. Está em absorver as ciências cognitivas, como a psicologia evolutiva e as heurísticas e vieses. Está em ler livros de história... “Estude ciência, não apenas eu!” é provavelmente o conselho mais importante que Ayn Rand deveria ter dado aos seus seguidores, mas não deu. Não há ser humano que já tenha vivido cujos ombros sejam largos o suficiente para carregar todo o peso de uma verdadeira ciência, a qual exige muitos colaboradores.

É interessante notar que os heróis fictícios de Ayn Rand eram arquitetos e engenheiros; John Galt, seu último herói, era um físico. No entanto, Ayn Rand ela mesma não era uma grande cientista. Pelo que sei, ela não era especialmente boa em matemática. Ela não poderia aspirar a rivalizar com seus próprios heróis. Talvez tenha sido por isso que ela tenha começado a perder a motivação para continuar se aprimorando.

Agora, eu, por exemplo, admiro a [audácia de Francis Bacon](#), mas ainda mantenho a humildade suficiente para admitir: “Se eu pudesse voltar no tempo e de alguma forma fazer Francis Bacon entender o problema com o qual [estou trabalhando atualmente](#), seus olhos saltariam das órbitas como rolhas de champanhe e explodiriam”.

Eu admiro as realizações de Newton. No entanto, minha postura em relação ao direito de voto das mulheres me impede de aceitar Newton como um modelo moral. Da mesma forma, meu conhecimento de probabilidade bayesiana me impede de considerar Newton como a autoridade suprema em conhecimento matemático. E meu conhecimento da Relatividade Especial, por mais limitado e pouco utilizado que seja, me impede de ver Newton como a máxima autoridade em física.

Realisticamente falando, Newton não poderia ter descoberto nenhuma das ideias que estou apresentando a ele — mas o progresso não é justo! E essa é a questão!

A ciência tem heróis, mas não deuses. Os grandes nomes não são nossos superiores, nem mesmo nossos rivais; eles são marcos que passamos em nossa jornada. E o marco mais importante é o herói que continua por vir.

Ser apenas mais um marco na estrada da humanidade é o melhor que se pode dizer sobre alguém, mas isso parecia muito humilde para agradar Ayn Rand. E foi assim que ela se tornou apenas uma Profeta Suprema.

## Referências

[1] Michael Shermer, "The Unlikeliest Cult in History," *Skeptic* 2, no. 2 (1993): 74–81, [http://www.2think.org/02\\_2\\_she.shtml](http://www.2think.org/02_2_she.shtml).

## 116 - Dois *Koans*<sup>52</sup> de Seita



Um racionalista novato, que estudava com o mestre *Ougi*<sup>53</sup>, foi repreendido por um amigo que disse: “Você passa todo esse tempo ouvindo seu mestre e falando sobre ‘racional’ isso e ‘racional’ aquilo — você entrou para uma seita!”

O noviço ficou profundamente perturbado; ele ouviu as palavras “Você entrou para uma seita!” ecoando em seus ouvidos enquanto estava deitado na cama naquela noite, e até mesmo em seus sonhos.

No dia seguinte, o noviço se aproximou de *Ougi* e relatou os acontecimentos, dizendo: “Mestre, estou constantemente preocupado que tudo isso seja realmente uma seita e que seus ensinamentos sejam apenas dogmas”.

*Ougi* respondeu: “Se você encontrar um martelo caído na estrada e resolver vendê-lo, poderá pedir um preço alto ou baixo. Mas se você ficar com o martelo e usá-lo para martelar pregos, quem poderá duvidar de seu valor?”

O noviço disse: “Mestre, é exatamente isso que me preocupa — suas respostas zen misteriosas”.

*Ougi* disse: “Tudo bem, então, falarei mais claramente e apresentarei argumentos perfeitamente razoáveis que demonstram que você não está em uma seita. Mas primeiro você precisa usar este chapéu bobo”.

*Ougi* deu ao novato um enorme chapéu de cowboy marrom de dez galões.

“Mestre...”, disse o noviço, hesitante.

“Quando eu tiver explicado tudo para você”, disse *Ougi*, “você entenderá por que isso foi necessário. Caso contrário, você pode continuar passando noites em claro, se perguntando se isto é uma seita”.

O noviço colocou o chapéu de cowboy.

*Ougi* disse: “Por quanto tempo você irá repetir minhas palavras e ignorar o significado? Pensamentos desordenados começam como apego a conclusões preferidas. Você está muito preocupado com sua imagem como racionalista. Você veio até mim em busca de reafirmação. Se você fosse [verdadeiramente curioso](#), sem saber de uma maneira ou de outra, teria pensado em formas de [resolver suas dúvidas](#). Como você precisava resolver sua dissonância cognitiva, estava disposto a usar um chapéu bobo. Se eu fosse uma pessoa má, poderia ter feito você pagar cem moedas de prata. Quando você se concentra em questões do mundo real, o valor ou a inutilidade do seu entendimento logo se tornará evidente. Você é como um espadachim que fica olhando para o lado para ver se alguém pode estar rindo dele...”

“Está tudo bem”, disse o noviço.

---

52 NT.: *Koans* são como enigmas zen-budistas que desafiam a lógica e provocam a intuição. São histórias, perguntas ou diálogos paradoxais, sem uma resposta única e «certa», que buscam romper o pensamento conceitual e levar o praticante a um entendimento mais profundo e direto da realidade. A palavra *koan* tem origem no chinês “*gōng'àn*”, que significa literalmente “caso público”, referindo-se a um precedente legal usado na antiga China. No Zen, o *koan* se tornou um “caso” para a mente investigar, um desafio para despertar a sabedoria inata.

53 NT.: *Ougi* (弱) do original japonês. Tradução aproximada em português brasileiro seria, “leque dobrável”, podendo evocar “leveza” e “delicadeza”.

“Você pediu a versão completa”, disse *Ougi*.

Mais tarde, esse novato sucedeu *Ougi* e ficou conhecido como *Ni no Tachi*<sup>54</sup>. A partir de então, ele não permitia que seus alunos citassem suas palavras durante debates, dizendo: “Utilizem as técnicas, mas não as mencionem”.

Um racionalista novato se aproximou do mestre *Ougi* e disse: “Mestre, estou preocupado de que nosso *dojo* da racionalidade possa ser... bem... um pouco sectário.”

“Isso é uma grande preocupação”, concordou *Ougi*.

O noviço esperou um momento, mas *Ougi* não disse mais nada.

Então o noviço falou novamente: “Quero dizer, desculpe-me, mas ter que usar essas vestes e esse capuz... parece que somos uma espécie de maçons estranhos ou algo assim.”

“Ah”, disse *Ougi*, “as vestimentas e adereços.”

“Bem, sim, as vestimentas e adereços”, confirmou o noviço. “Parece extremamente irracional.”

“Abordarei todas as suas preocupações”, disse o mestre, “mas primeiro você precisa colocar este chapéu bobo.” E *Ougi* lhe apresentou um chapéu de mago bordado com luas e estrelas.

O noviço pegou o chapéu, olhou para ele e em seguida explodiu de frustração: “Como isso pode ajudar?”

“Já que você está tão preocupado com as interações das vestimentas com a teoria das probabilidades”, explicou *Ougi*, “não deveria surpreendê-lo que seja necessário usar um chapéu especial para entender.”

Quando o noviço alcançou o status de aluno de pós-graduação, adotou o nome de *Bouzo*<sup>55</sup> e só discutia racionalidade enquanto vestia uma roupa de palhaço.

---

54 NT.: A tradução literal de *Ni no Tachi* (二の太刀) do japonês para o português é “**Segundo Corte**” ou “**Segundo Golpe**”. O *Ni no Tachi* é um conceito fundamental nas artes marciais japonesas, representando o segundo golpe em uma sequência de ataques. É um movimento estratégico crucial, frequentemente executado para explorar a vulnerabilidade do oponente após o primeiro golpe ou para contra-atacar durante sua recuperação. A maestria do *Ni no Tachi* exige timing preciso, leitura dos movimentos do adversário e a capacidade de capitalizar sobre a abertura criada pelo primeiro ataque.

55 NT.: *Bouzo* (坊主), palavra de origem japonesa, pode ser traduzida livremente como “monge” ou “cabeça raspada”. Refere-se especificamente aos monges budistas que raspam a cabeça como parte de sua prática religiosa.

## 117 - Experimento de Conformidade de Asch



Solomon Asch, por meio de experimentos realizados originalmente na década de 1950 e replicados com sucesso desde então, identificou um fenômeno conhecido como “conformidade”. No experimento clássico, um participante é apresentado a um quebra-cabeça com o seguinte diagrama: qual das linhas A, B e C tem o mesmo tamanho da linha X? Tire um momento para determinar sua própria resposta...



A pegadinha é que o participante está sentado ao lado de várias outras pessoas que também estão examinando o diagrama — aparentemente outros participantes, mas na realidade, são cúmplices do experimentador. Os outros “participantes” do experimento, um após o outro, afirmam que a linha C parece ter o mesmo tamanho que X. O verdadeiro participante está sentado na penúltima posição. Quantas pessoas, nessas circunstâncias, diriam “C” — fornecendo uma resposta obviamente incorreta que concorda com a resposta unânime dos outros participantes? Qual você acha que seria a porcentagem?

Três quartos dos participantes do experimento de Asch deram uma resposta “conforme” pelo menos uma vez. Um terço dos participantes se conformou mais da metade do tempo.

Entrevistas realizadas após o experimento mostraram que, embora a maioria dos participantes afirmasse não ter realmente acreditado em suas respostas “conformes”, alguns disseram que realmente achavam que a opção “conforme” era a correta.

Asch ficou perturbado com esses resultados:

O fato de encontrarmos uma tendência tão forte à conformidade em nossa sociedade... é uma questão preocupante. Isso levanta questões sobre nossos métodos educacionais e os valores que norteiam nosso comportamento<sup>56</sup> [1].

Não é uma questão trivial determinar se os participantes dos experimentos de Asch agiram irra-

---

56 NT.: Texto original em inglês. *That we have found the tendency to conformity in our society so strong... is a matter of concern. It raises questions about our ways of education and about the values that guide our conduct.*

cionalmente. O Teorema da Concordância de Robert Aumann demonstra que bayesianos honestos não podem concordar em discordar — se eles têm conhecimento comum de suas estimativas de probabilidade, eles têm a mesma estimativa de probabilidade. O Teorema da Concordância de Aumann foi comprovado mais de vinte anos após os experimentos de Asch, mas apenas formaliza e fortalece um ponto intuitivamente óbvio — as crenças de outras pessoas costumam ser evidências legítimas.

Em um experimento, suponha que você esteja olhando para um diagrama como o mencionado acima e saiba que todos os outros participantes são honestos e estão vendo o mesmo diagrama que você. Se três outras pessoas disserem que C é do mesmo tamanho que X, quais são as chances de você ser o único correto?

Não estou reivindicando nenhuma vantagem em termos de percepção visual. Não acredito que seja melhor do que uma pessoa comum em julgar se duas linhas têm o mesmo tamanho. Em termos de racionalidade individual, reconheço minha própria confusão significativa e, portanto, atribuo uma probabilidade maior que 50% ao voto da maioria.

Em termos de racionalidade de grupo, parece-me que a coisa certa a dizer como um racionalista honesto seria: “É surpreendente, parece-me que B tem o mesmo tamanho que X. No entanto, se todos estivermos olhando para o mesmo diagrama e relatando honestamente, não tenho motivo para acreditar que minha avaliação seja melhor do que a sua.” A última frase é importante — é uma alegação de desacordo muito mais sutil do que dizer: “Ah, estou percebendo a ilusão de ótica — entendo por que você acha que é C, é claro, mas a resposta correta é B”.

Portanto, os participantes que se conformam nesses experimentos não são automaticamente condenados por irracionalidade, com base no que foi descrito até agora. No entanto, como você pode esperar, o diabo está nos detalhes dos resultados experimentais. De acordo com uma meta-análise de mais de cem replicações conduzida por Smith e Bond [2]:

A conformidade aumenta significativamente até 3 cúmplices, mas não aumenta ainda mais com a presença de 10 a 15 cúmplices. Se as pessoas estão se conformando de maneira racional, então a opinião de 15 outros participantes deveria ser uma evidência substancialmente mais forte do que a opinião de apenas 3.

A adição de um único dissidente — apenas mais uma pessoa que dá a resposta correta, ou mesmo uma resposta incorreta diferente da resposta incorreta do grupo — reduz a conformidade acentuadamente, para cerca de 5 a 10% dos participantes. Se você estiver aplicando alguma versão intuitiva do Acordo de Aumann e pensando que quando 1 pessoa discorda de 3, as 3 provavelmente estão corretas, então, geralmente, você deve estar igualmente disposto a considerar que 2 pessoas discordariam de 6 pessoas (Não é automaticamente verdade, mas verdade *ceteris paribus* — todas as outras coisas sendo iguais). Por outro lado, se você estiver lidando com pessoas emocionalmente desconfortáveis por serem diferentes, é fácil ver como uma única pessoa que concorda com você, ou mesmo uma única pessoa que discorda do grupo, diminuiria consideravelmente o desconforto.

Não surpreendentemente, participantes na condição de dissidente não acreditavam que sua não conformidade fosse influenciada ou possibilitada pelo dissidente. Assim como os 90% dos motoristas que acreditam estar ~~acima da~~ na média entre os 50% melhores, alguns deles podem estar corretos, mas não todos. As pessoas não têm consciência das causas de sua conformidade ou dissidência, enfraquecendo a tentativa de argumentar a favor de sua manifestação como resultado de racionalidade. Por exemplo, na hipótese de que as pessoas estão escolhendo racionalmente mentir para não se destacarem, parece que alguns participantes na condição de dissidente não antecipam conscientemente a “estratégia consciente” que empregariam quando confrontados com oposição unânime.

Quando o único dissidente de repente passou a se conformar com o grupo, as taxas de conformidade dos participantes voltaram a ser tão altas quanto na condição em que não havia dissidente. Ser o primeiro dissidente é um serviço social valioso (e custoso!), mas você precisa continuar sendo.

Consistente tanto dentro quanto entre os experimentos, grupos exclusivamente femininos (com uma participante feminina ao lado de cúmplices femininas) se conformam significativamente com mais frequência do que grupos exclusivamente masculinos. Cerca de metade das mulheres se conforma mais da metade do tempo, em comparação com um terço dos homens. Se argumentarmos que o participante médio é racional, então aparentemente as mulheres são muito agradáveis e os homens são muito desagradáveis, portanto,



nenhum dos grupos é verdadeiramente racional...

Manipulações com grupos internos e externos (por exemplo, um participante deficiente ao lado de outros participantes deficientes) também mostram que a conformidade é significativamente maior entre os membros de um grupo interno.

A conformidade é menor quando se trata de diagramas óbvios, como o mencionado no início deste ensaio, em comparação com diagramas nos quais os erros são mais sutis. Isso é difícil de explicar se todos os participantes estão tomando uma decisão socialmente racional para evitar se destacar.

Paul Crowley me lembra de observar que quando os participantes podem responder de uma maneira que não será vista pelo grupo, a conformidade também diminui, o que também contradiz uma interpretação de Aumann.

## Referências

[1] Solomon E. Asch, "Studies of Independence and Conformity: A Minority of One Against a Unanimous Majority," *Psychological Monographs* 70 (1956).

[2] Rod Bond and Peter B. Smith, "Culture and Conformity: A Meta-Analysis of Studies Using Asch's (1952b, 1956) Line Judgment Task," *Psychological Bulletin* 119 (1996): 111–137.

## 118 - Sobre Como Expressar Suas Preocupações



O aspecto mais perturbador dos [experimentos de conformidade de Asch](#) é que você pode fazer com que muitas pessoas afirmem que preto é branco, se as colocar em uma sala cheia de outras pessoas dizendo o mesmo.

O que é esperançoso nos experimentos de conformidade de Asch é que um único dissidente reduziu significativamente a taxa de conformidade, mesmo que o dissidente estivesse apenas dando uma resposta errada diferente. E o mais surpreendente é que a dissidência não foi aprendida ao longo do experimento: quando o único dissidente começou a concordar com o grupo, as taxas de conformidade voltaram a aumentar.

Ser uma voz dissidente pode trazer benefícios reais para o grupo. Mas também (notoriamente) tem um custo. E então você precisa mantê-lo. Além disso, é possível que você esteja errado.

Recentemente, tive uma experiência interessante em que comecei a discutir um projeto com duas pessoas que já haviam feito algum planejamento por conta própria. Eu achei que eles estavam sendo muito otimistas e fiz várias sugestões que envolviam uma margem de segurança para o projeto. Logo, um quarto indivíduo apareceu, que estava dando carona para um dos outros dois, e começou a fazer sugestões. Nesse momento, tive uma percepção repentina de como os grupos se tornam excessivamente confiantes, porque toda vez que eu levantava uma possível preocupação, o quarto indivíduo dizia: “Não se preocupe, tenho certeza de que podemos lidar com isso!” ou algo igualmente reconfortante.

Um indivíduo, trabalhando sozinho, naturalmente terá dúvidas. Eles irão pensar consigo mesmos: “Será que realmente consigo fazer XYZ?” porque não há nada de errado em duvidar de sua própria competência. Mas quando duas pessoas inseguras formam um grupo, é considerado educado dizer coisas agradáveis e tranquilizadoras, enquanto é considerado indelicado questionar a competência do outro. Juntos, eles se tornam mais otimistas do que seriam individualmente, com as dúvidas de cada um sendo suprimidas pela aparente segurança do outro, sem perceberem que o outro indivíduo tinha inicialmente as mesmas dúvidas internas.

A possibilidade mais assustadora levantada pelos experimentos de conformidade de Asch é a de todos concordarem com o grupo, influenciados pelas vozes confiantes dos outros, tendo cuidado para não demonstrar suas próprias dúvidas, sem perceber que os outros estão suprimindo preocupações semelhantes. Isso é conhecido como “ignorância pluralista”.

Robin Hanson e eu tivemos um longo debate sobre quando os aspirantes a racionalistas devem se atrever a discordar.

Eu tendo a adotar a posição amplamente defendida de que você não tem escolha real além de formar suas próprias opiniões.

Robin Hanson defende uma posição mais iconoclasta, afirmando que você — não apenas outras pessoas — deve considerar que os outros podem ser mais sábios. Independentemente de nossas várias disputas, concordamos que o Teorema de Acordo de Aumann se estende para implicar que o conhecimento comum de um desacordo factual mostra que alguém deve estar sendo [irracional](#). Apesar dos olhares estranhos que recebemos, permanecemos firmes em nossa posição sobre a modéstia:

“Esqueça o que todos lhe dizem sobre individualismo, você deve prestar atenção ao que as outras pessoas pensam.”

A questão é que, para os racionalistas, discordar do grupo é algo sério. Não se pode simplesmente acenar com o argumento de [“Todo mundo tem direito à sua própria opinião”](#).

Acredito que a lição mais importante a ser extraída dos experimentos de Asch é distinguir entre “expressar preocupação” e “discordar”. Levantar um ponto que os outros não tenham abordado não significa necessariamente discordar do grupo no final da discussão.

O processo de convergência do ideal bayesiano envolve compartilhar evidências imprevisíveis para o ouvinte. O resultado do acordo de Aumann só se aplica ao conhecimento comum, onde você sabe, eu sei, você sabe que eu sei, etc. O post ou artigo de Hanson sobre o assunto, intitulado [“Não Podemos Prever Discordância”](#) fornece uma imagem de quão estranho pareceria observar racionalistas ideais convergindo em uma estimativa de probabilidade; não se parece em nada com dois negociadores em um mercado convergindo para um preço.

Infelizmente, não há muita diferença social entre “expressar preocupações” e “discordância”. Um grupo de racionalistas pode concordar em fingir que existe uma diferença, mas não é assim que os seres humanos realmente funcionam. Assim que você se manifesta, você cometeu um ato socialmente irrevogável; você se tornou o prego que sobressai, a discórdia na harmonia confortável do grupo e você não pode desfazer isso. Qualquer pessoa ofendida por uma preocupação que você expressou sobre a competência deles para concluir a tarefa XYZ, provavelmente guardará rancor mesmo se você disser “Sem problemas, vou acompanhar o grupo” no final.

O experimento de Asch mostra que o poder da dissidência para inspirar os outros é real. O experimento de Asch mostra que o poder da conformidade é real. Se todos se absterem de expressar suas dúvidas privadas, isso realmente levará os grupos à loucura. Mas a história está repleta de lições sobre o preço de ser o primeiro, ou mesmo o segundo, a dizer que o Imperador está nu. Nem as pessoas estão programadas para distinguir “expressar uma preocupação” de “discordância, mesmo com conhecimento comum”; essa distinção é uma artificialidade dos racionalistas. Se você ler os livros de autoajuda mais cínicos (por exemplo, O Príncipe, de Maquiavel), eles lhe aconselharão a mascarar completamente sua não conformidade, não expressar suas preocupações primeiro e depois concordar no final. Se você prestar o serviço ao grupo de ser aquele que dá voz aos problemas óbvios, não espere que o grupo lhe agradeça por isso.

Esses são os custos e os benefícios da dissidência — se você “discorda” ou apenas “expressa preocupação” — e a decisão cabe a você.

## 119 - Dissidência solitária



O [experimento de conformidade de Asch](#) mostrou que a presença de um único dissidente reduzia tremendamente a incidência de respostas erradas “conformes”. O individualismo é fácil, mostra a experiência, quando você tem companhia em seu desafio. Todos os outros sujeitos na sala, exceto um, dizem que preto é branco. Você se torna a segunda pessoa a dizer que preto é preto. E é glorioso: vocês dois, rebeldes solitários e desafiadores, contra o mundo! As entrevistas posteriores mostraram que os sujeitos, quando havia um dissidente presente, expressaram fortes sentimentos de camaradagem com esse dissidente — embora, é claro, não achassem que a presença do dissidente tivesse influenciado sua própria inconformidade.

Mas você só pode se juntar à rebelião depois que alguém, em algum lugar, se tornar o primeiro a se rebelar. Alguém tem que dizer que preto é preto após ouvir todos os outros, um após o outro, dizerem que preto é branco. E isso — mostra a experiência — é muito mais difícil.

A dissidência solitária não é como ir para a escola [vestida de preto](#). É como ir para a escola vestindo uma roupa de palhaço.

Há uma grande diferença entre entrar para a rebelião e deixar o bando.

Se há uma coisa que não suporto, é falsidade — você já deve ter percebido isso. Bem, a dissidência solitária deve ser uma das características mais comumente e ostensivamente falsificadas por aí. Todo mundo quer ser um iconoclasta.

Não pretendo diminuir o ato de ingressar em uma rebelião. Há rebeliões que vale a pena aderir. É preciso coragem para enfrentar a desaprovação de seu grupo de colegas, ou talvez ainda pior, seu encolher de ombros. Desnecessário dizer que ir a um show de rock não é rebeldia. Mas, por exemplo, o vegetarianismo é. Eu mesmo não sou vegetariano, mas respeito as pessoas que são, porque acho que é preciso uma quantidade notável de coragem silenciosa para dizer às pessoas que hambúrgueres não funcionam no jantar. Embora na Bay Area as pessoas perguntem por hábito.

Ainda assim, se você disser às pessoas que é vegetariano, elas pensarão que entendem seus motivos (mesmo que não entendam). Elas podem discordar. Elas podem ficar ofendidas se você anunciar sua escolha com orgulho excessivo ou, nesse caso, podem ficar ofendidas apenas porque se ofendem facilmente. Mas elas sabem como se relacionar com você.

Quando alguém se veste de preto para ir à escola, os professores e as outras crianças entendem o papel que está sendo assumido em sua sociedade. Isso está fora do sistema - de uma forma muito padronizada que todos reconhecem e entendem. Não, sabe, realmente fora do sistema. É um desafio ao pensamento padrão, de um tipo padrão, para as pessoas dizerem indignadas: “Não consigo entender por que você...”, mas não precisem realmente ter pensamentos que não haviam pensado antes. Como diz o ditado: “Alguma das ‘literaturas subversivas’ que você leu fez com que você modificasse alguma de suas visões políticas?”

O que realmente exige coragem é enfrentar a total incompreensão das pessoas ao seu redor, quando você faz algo que não é Rebelião Padrão nº 37, algo para o qual eles não têm um roteiro pronto. Elas não te odeiam por ser um rebelde. Elas apenas acham que você é estranho e se afastam. Essa perspectiva gera um medo muito mais profundo. É a diferença entre explicar o vegetarianismo e explicar a [criogenia](#). Existem outros adeptos da criogenia no mundo, em algum lugar, mas eles não estão perto de você. Você tem que explicar, sozinho, para as pessoas que simplesmente acham estranho. Não é proibido, mas está fora dos limites que as

peças nem pensam. Você vai ficar com a cabeça congelada? Você acha que isso impedirá que você morra? O que quer dizer com informações cerebrais? Hein? O quê? Você está louco?

Sou tentado a ensaiar uma explicação pós-fato na psicologia evolutiva: você poderia se reunir com um pequeno grupo de amigos e se afastar de seu bando de caçadores-coletores, mas ter que ir sozinho na floresta provavelmente era uma sentença de morte — pelo menos reprodutivamente. Não raciocinamos sobre isso explicitamente, mas essa não é a natureza da psicologia evolutiva. Juntar-se a uma rebelião que todo mundo conhece é assustador, mas nem de longe tão assustador quanto fazer algo realmente diferente — algo que em tempos ancestrais poderia ter resultado, não na separação da banda, mas em você sendo expulso sozinho.

Como testemunha o caso da criogenia, o medo de pensar realmente diferente é mais forte que o medo da morte. Os caçadores-coletores precisavam estar preparados para enfrentar a morte rotineiramente — caçando grandes mamíferos ou apenas caminhando em um mundo que continha predadores. Eles precisavam dessa coragem para viver. Coragem para desafiar os modos de pensar padrão da tribo, para entreter pensamentos que parecem realmente estranhos - bem, isso provavelmente não serviu bem aos seus portadores. Não raciocinamos sobre isso explicitamente; não é assim que a psicologia evolutiva funciona. Nós, seres humanos, somos construídos de tal maneira que muitos de nós saltamos de paraquedas em vez de nos inscrevermos na criogenia.

E isso nem é a maior coragem. Há mais de um adepto da criogenia no mundo. Apenas Robert Ettinger teve que dizer isso primeiro.

Para ser um revolucionário científico, você precisa ser a primeira pessoa a contradizer o que todo mundo que você conhece está pensando. Este não é o único caminho para a grandeza científica; é raro mesmo entre os grandes. Ninguém pode se tornar um revolucionário científico tentando imitar o revolucionarismo. Você só pode chegar lá buscando a resposta correta em todas as coisas, seja a resposta correta revolucionária ou não. Mas se, no decorrer do tempo - se, tendo absorvido todo o poder e sabedoria do conhecimento que já acumulou - se, depois de tudo isso e de uma dose de pura sorte, você descobrir que sua busca pela mera correção o levará a novo território... então você tem uma oportunidade para sua coragem falhar.

Essa é a verdadeira coragem da dissidência solitária, aquela que toda banda de rock por aí tenta imitar.

Claro, nem tudo que exige coragem é uma boa ideia. Seria preciso coragem para cair de um penhasco, mas você simplesmente cairia.

O receio de ser dissidente solitário pode ser um obstáculo para o surgimento de boas ideias. No entanto, nem toda ideia divergente é boa. (Veja também [Against Free Thinkers](#) (Contra livres pensadores) de Robin Hanson). A maior parte da dificuldade em produzir uma nova ideia científica está justamente na parte “científica” dela ser verdadeira.

Realmente não é necessário ser diferente por ser diferente. Se você fizer as coisas de maneira diferente apenas quando vir uma razão esmagadoramente boa, terá problemas mais do que suficientes para durar o resto de sua vida.

Existem alguns grupos genuínos de iconoclastas por aí. A Igreja do SubGenius, por exemplo, parece genuinamente visar confundir os mundanos, não apenas os ofender. E há ilhas de tolerância genuína no mundo, como as convenções de ficção científica. Existem certas pessoas que não têm medo de deixar o bando. Muito menos pessoas realmente existem do que se imaginam rebeldes; mas eles existem. No entanto, os revolucionários científicos são tremendamente mais raros. Reflita sobre isso.

Quanto a mim, bem, eu realmente sou um iconoclasta. Todo mundo pensa que é, mas comigo é verdade, entende? Eu teria *super* vestido uma roupa de palhaço para a escola. Minhas conversas sérias eram com livros, não com outras crianças.

Mas se você acha que *super* usaria aquela roupa de palhaço, também não se orgulhe disso! Significa apenas que você precisa fazer um esforço na direção oposta para evitar discordar com muita facilidade. Isso é o que eu tenho que fazer, para corrigir minha própria natureza. Outras pessoas têm razões para pensar o

que pensam e ignorar isso completamente é tão ruim quanto ter medo de contradizê-las. Você não gostaria de acabar como um pensador livre. Não é uma virtude, veja você — apenas um viés de qualquer maneira.

## 120 - Anti-sectarismo sectário



No mundo moderno, entrar para uma seita é provavelmente uma das piores coisas que podem acontecer com alguém. No melhor dos casos, você acaba em um grupo de pessoas sinceras, mas iludidas, cometendo um erro honesto, porém bem-comportado, e gastando muito tempo e dinheiro sem obter nada em troca. Na verdade, isso poderia descrever qualquer startup fracassada do Vale do Silício. É uma experiência angustiante, se pararmos para pensar. Então, sim, é bastante assustador.

Mas as verdadeiras seitas são muito piores. Eles usam técnicas de recrutamento como o “bombardeio de amor”, direcionado a pessoas em crise pessoal. Privação de sono. Exaustão causada por trabalho árduo. Comunidades isoladas para separar os recrutas de amigos e familiares. Reuniões diárias para confessar pensamentos impuros. Não é incomum que as seitas tomem todo o dinheiro dos recrutas — economias de uma vida inteira e o salário semanal, forçando-os a depender da seita para comida e roupas. A fome é usada como punição pela desobediência. Há uma lavagem cerebral intensa e danos graves.

Considerando tudo isso, talvez eu devesse ter mais empatia por pessoas terrivelmente nervosas ao se envolverem em algum empreendimento aparentemente estranho, que talvez podem estar se juntando a uma seita. Não deveria me irritar tanto. Mas acontece.

Ponto número um: “Seitas” e “não-seitas” não são categorias naturais separadas, como cães e gatos. Se você olhar para qualquer lista de características de seitas, verá itens que poderiam facilmente descrever partidos políticos e corporações — “membros do grupo encorajados a desconfiar de críticas externas por terem motivos ocultos”, “estrutura hierárquica autoritária”. Eu já escrevi sobre falhas em grupos, como a [polarização](#), [espirais de autodestruição feliz](#), [falta de criticidade](#) e [resfriamento evaporativo](#), todos os quais parecem se alimentar uns dos outros. Quando essas falhas giram juntas e se encontram, elas se combinam para formar uma Super-Falha mais estúpida do que qualquer uma das partes, como Voltron. Mas isso não é uma essência de seita; é um atrator de seita.

Os cães nascem com DNA canino e os gatos nascem com DNA felino. No mundo atual, não há meio-termo. (Mesmo com manipulação genética, não seria tão simples criar um organismo com genes meio caninos e meio felinos.) Não é como se houvesse um conjunto de características caninas que se reforçam mutuamente, onde um gato individual poderia entrar e se tornar meio cão.

A mente humana, ao pensar em categorias, parece preferir essências a atratores. As pessoas querem dizer: “Isso é uma seita” ou “Isso não é uma seita”, e então a tarefa de classificação está concluída. Se você observar que Sócrates tem dez dedos, usa roupas e fala fluentemente grego, então você pode dizer “Sócrates é humano” e deduzir “Sócrates é vulnerável à cicuta” sem fazer testes específicos de sangue para confirmar sua mortalidade. Você decidiu sobre a humanidade de Sócrates de uma vez por todas.

Ao analisar um determinado grupo de indivíduos, é essencial considerar a possibilidade de polarização interna e um [efeito halo](#) positivo em torno de sua paixão — seja [objetivismo](#), vegetarianismo ou [redes neurais](#). No entanto, a presença desses fenômenos não permite, com base nas evidências disponíveis até o momento, concluir que o grupo apresenta falta de criticidade.

Você não pode deduzir se a principal ideia deles é verdadeira, falsa ou genuinamente útil, mas não tão útil quanto eles pensam. Pelas informações coletadas até o momento, você não pode deduzir se eles são educados ou se o atrairão para o isolamento e privação de sono e comida. As características de seita não estão todas presentes ou ausentes.

Se você examinar argumentos online sobre “X é uma seita” versus “X não é uma seita”, verá que um lado usa uma lista online de características de seita e encontra uma que se aplica, concluindo “Portanto, é uma seita!”. Enquanto o outro lado encontra uma característica que não se aplica, concluindo “Portanto, não é uma seita!”.

Você não pode obter uma imagem precisa da dinâmica de raciocínio de um grupo usando esse tipo de essencialismo. É necessário prestar atenção às características individuais individualmente.

Além disso, [estupidez ao contrário não é inteligência](#). Se você está interessado na ideia central, não apenas no grupo de implementação, ideias inteligentes podem ter seguidores estúpidos. Muitos adeptos da Nova Era falam sobre “física quântica”, mas isso não é uma crítica à física quântica. Claro, ideias estúpidas também podem ter seguidores estúpidos. Com o essencialismo binário vem a ideia de que se você inferir que um grupo é uma “seita”, então suas crenças devem ser falsas, porque as falsas crenças são características das seitas, assim como os gatos têm pelos. Se você está interessado na ideia, [olhe para a ideia, não para as pessoas](#). Sectarismo é uma característica de grupos, não de hipóteses.

O segundo erro é quando as pessoas perguntam nervosamente: “Isso não é uma seita, é?” Parece-me que estão buscando a garantia da racionalidade. A noção de um racionalista não se apegando demais à sua autoimagem como racionalista merece seu próprio ensaio (consulte Doze Virtudes, Por que verdade? E... e [Dois koans de seita](#)). Mas mesmo sem entrar em detalhes, certamente pode-se ver que a busca nervosa por segurança não é o melhor estado de espírito para avaliar questões de racionalidade. Você não ficará [genuinamente curioso](#) ou pensará em maneiras de [sanar suas dúvidas](#). Em vez disso, você encontrará alguma fonte online que diz que seitas usam privação do sono para controlar as pessoas, notará que seu grupo favorito não usa privação do sono e concluirá: “Não é uma seita. Ufa!” Se não tem pelos, não deve ser um gato. Muito reconfortante.

Mas [toda causa quer ser uma seita](#), seja a causa sábia ou tola. A dicotomia in-grupos e ex-grupos faz parte da natureza humana, não uma [maldição especial](#) dos [mutantes](#). A racionalidade é a exceção, não, a regra. É necessário um esforço contínuo para manter a racionalidade contra o deslizamento natural em direção à entropia. Se você decidir “Não é uma seita!” e suspirar aliviado, não fará um esforço constante para combater as tendências comuns das seitas. Você decidirá que a essência de seita está ausente e parará de combater a entropia do atrator da seita.

Se você está terrivelmente nervoso com a seita, então você vai querer negar qualquer indício de qualquer característica que se assemelhe a uma seita. Mas qualquer grupo com um objetivo visto positivamente arrisca ser afetado pelo efeito halo e terá que combater a entropia para evitar uma espiral de morte afetiva. Isso é verdade até mesmo para instituições comuns como partidos políticos - pessoas que pensam que “valores liberais” ou “valores conservadores” podem curar o câncer, por exemplo. É verdade para startups do Vale do Silício, tanto as fracassadas quanto as bem-sucedidas. É verdade para usuários de Mac e para usuários de Linux. O efeito halo não se torna bom só porque todo mundo faz isso; se todo mundo cair de um penhasco, você também não deve fazer isso. O erro de raciocínio deve ser combatido, não tolerado. Mas se você está muito nervoso com “Tem certeza de que isso não é uma seita?”, então você ficará relutante em ver qualquer sinal de seita, porque isso implicaria que você está em uma seita, e não é uma seita! Portanto, você não perceberá os campos de batalha atuais, onde as tendências comuns a seita estão avançando ou sendo repelidas.

O terceiro erro em perguntar nervosamente “Isso não é uma seita, é?” É que, suspeito fortemente, o nervosismo existe pelas razões totalmente erradas.

Porque os grupos que exaltam sua “Coisa Feliz” até as estrelas, encorajam os membros a doar todo o seu dinheiro e trabalhar em servidão voluntária, e administram complexos privados onde os membros são mantidos rigorosamente isolados, são chamados de “religiões” em vez de “seitas”. Eles existem há algumas centenas de anos?

Por que a maioria das pessoas que faz perguntas nervosas sobre criogenia e pergunta: “Isso não é uma seita, é?” não ficariam igualmente nervosas em participar de um comício político republicano ou democrata? Dicotomias entre grupo interno e externo e espirais de morte feliz podem ocorrer em discussões políticas, religiões dominantes e até mesmo no *fandom* de esportes. Se o nervosismo surgisse do medo de



erros de racionalidade, as pessoas perguntariam: “Isso não é uma dicotomia de grupo interno e externo?” sobre comícios políticos democratas ou republicanos, com o mesmo medo em sua voz.

Existe uma razão legítima para ter menos medo do libertarianismo do que de uma seita ao disco voador, porque os libertários não têm a reputação de usar privação do sono para converter pessoas. No entanto, os criogenistas também não têm essa reputação. Então, por que ficar mais preocupado com [o congelamento da sua cabeça após parar de respirar?](#)

Suspeito que o nervosismo não seja o medo de acreditar em algo falso ou o medo de danos físicos. É o medo da [dissidência solitária](#). É a sensação nervosa que os sujeitos experimentam no famoso experimento de [conformidade de Asch](#), quando todos os outros sujeitos (na verdade, cúmplices) afirmam repetidamente que a linha C tem o mesmo tamanho da linha X, fazendo com que a linha B pareça do mesmo tamanho da linha X para o sujeito. É o medo de se separar do grupo.

É por isso que grupos cujas crenças existem há tempo suficiente para parecerem “normais” não inspiram o mesmo nervosismo que as “seitas”, mesmo que algumas religiões tradicionais também possam exigir que os fiéis doem todo o seu dinheiro para um mosteiro. É por isso que grupos como partidos políticos, frequentemente envolvidos em erros de racionalidade, não causam o mesmo nervosismo que as “seitas”. A palavra “seita” não está sendo usada aqui para representar erros de racionalidade, mas sim como um rótulo para algo que parece estranho.

Não toda mudança é uma melhoria, mas toda melhoria é necessariamente uma mudança. Se você deseja fazer algo melhor, não tem escolha a não ser fazer de forma diferente. A sabedoria comum contém uma quantidade significativa de, bem, sabedoria verdadeira. Portanto, faz sentido exigir uma carga extra de prova para aceitar o que é considerado estranho. No entanto, o nervosismo não é uma consideração deliberada e racional desse tipo. É o medo de acreditar em algo que fará com que seus amigos olhem para você estranhamente. É por isso que as pessoas perguntam “Isso não é uma seita, é?” em um tom que nunca usariam para perguntar se alguém vai a um comício político ou organiza um grande evento de Natal.

Isso é o que me incomoda.

É como se, assim que você passasse a acreditar em algo que seus ancestrais não acreditaram, a Fada da Seita descesse dos céus e infundisse em você a Essência da Seita, e a próxima coisa que você sabe é que todos estão usando túnicas e [cantando em coro](#). Como se as crenças “estranhas” fossem a causa direta dos problemas, independentemente da privação do sono e dos espancamentos. Os danos causados pelas seitas - como o suicídio em massa de Heaven’s Gate e outros - apenas reforçam a ideia de que todos com crenças estranhas são loucos; a característica primordial dos “membros de seita” é serem estranhos com comportamentos peculiares.

Sim, crenças socialmente incomuns colocam um grupo em risco de pensamento interno e externo, esgotamento evitativo e outros problemas. No entanto, a raridade em si é um fator de risco, não uma doença. O mesmo ocorre com ter um objetivo que você considera valioso. Independentemente de a crença ser verdadeira ou não, ter um objetivo nobre sempre coloca você em risco de cair na espiral da autodestruição feliz. Mas isso não torna os objetivos elevados uma doença, e sim um fator de risco. Alguns objetivos [valem realmente a pena perseguir](#).

Por outro lado, não vejo nenhuma razão legítima para a privação do sono ou para ameaçar dissidentes com espancamentos. Quando um grupo faz isso, seja chamado de “seita” ou “não-seita”, você respondeu diretamente à pergunta pragmática de se juntar ou não.

Problema quatro: o medo da dissidência solitária é algo que as próprias seitas exploram. O medo de que seus amigos o vejam com desaprovação é exatamente o efeito que as verdadeiras seitas usam para converter e manter os membros - cercar os convertidos com total concordância entre os crentes da seita.

O medo de ideias estranhas e o impulso para a conformidade certamente afastaram muitas vítimas em potencial das seitas aos discos voadores. Quando você está fora, isso o mantém afastado. Mas quando você está dentro, isso o mantém preso. A conformidade apenas o prende, não importa onde você esteja, seja em um lugar bom ou ruim.

Alguém desejaria que houvesse uma maneira de ter certeza de que não está em uma “seita”. Alguma resposta definitiva e contundente para aqueles que olham para você estranhamente. Alguma forma de saber com certeza se está fazendo a coisa certa, sem essas dúvidas constantes. Isso é o que chamamos de “necessidade de encerramento”. E, é claro, as seitas também exploram isso.

Daí a expressão “anti-sectarismo sectário”.

Viver com dúvidas não é uma virtude — o [propósito de toda dúvida é se aniquilar](#) no sucesso ou no fracasso, e uma dúvida que apenas paira não leva a lugar algum. Mas às vezes, uma dúvida demora a se dissipar. Viver com uma pilha de dúvidas atualmente não resolvidas é um fato inevitável da vida para os racionalistas. A dúvida não deve ser assustadora. Caso contrário, você terá que escolher entre levar uma vida de busca constante ou uma vida tola.

Se você realmente não consegue descobrir se um grupo é uma “seita”, então você terá que decidir em condições de incerteza. É disso que trata a teoria da decisão.

Problema cinco: falta de pensamento estratégico.

Conheço pessoas cautelosas em relação ao [singularismo](#), assim como são cautelosas em relação a partidos políticos e religiões tradicionais. Elas são cautelosas, não nervosas ou defensivas. Essas pessoas podem perceber rapidamente que o singularismo obviamente não é uma seita completa, com privação do sono e assim por diante, um agente definido principalmente por concordar com qualquer coisa boa que se diga sobre ele. Só porque algo não é uma seita agora, não significa que não possa se tornar um no futuro. Ser “sectário” é um atrator, não uma essência.

Esse tipo de cautela me incomoda? De forma alguma. Eu mesmo passo muito tempo preocupado com esse cenário. Tento antecipar minhas jogadas de Go para bloquear o movimento nessa direção. Por exemplo, é por isso que escrevo uma série de ensaios sobre falhas no pensamento sectário.

As pessoas que falam sobre “racionalidade” também possuem um fator de risco adicional. Dar conselhos sobre como pensar é intrinsecamente perigoso. No entanto, é um fator de risco, não uma doença.

Ambas as minhas causas favoritas estão em risco de se tornarem seitas. No entanto, por alguma razão, sou questionado com muito mais frequência: “Tem certeza de que isso não é uma seita?” quando falo sobre inteligências artificiais poderosas do que quando falo sobre teoria da probabilidade e ciência cognitiva. Não sei se um fator de risco é maior que o outro, mas sei qual soa mais estranho...

O problema número 6 de se perguntar “Isso não é uma seita, é?”...

A própria pergunta me coloca em um tipo irritante de dilema. Um verdadeiro guru do mal usaria certamente a ansiedade de alguém contra eles e criaria um argumento elaborado e plausível explicando por que isso não é uma seita, e a pessoa ficaria ansiosa para aceitá-lo. Às vezes, tenho a impressão de que é isso que as pessoas querem que eu faça! Sempre que tento escrever sobre seitas e como evitá-las, continuo sentindo que estou cedendo a esse desejo imperfeito — que estou, no final das contas, oferecendo segurança às pessoas. Mesmo quando digo ser necessário lutar constantemente contra a entropia.

Parece que estou me tornando o primeiro dissidente no experimento de conformidade de Asch, dizendo às pessoas: “Sim, a linha X é realmente igual à linha B, está tudo bem que você também diga isso”. Elas não deveriam precisar perguntar! Ou pior ainda, parece que estou apresentando um argumento elaborado sobre por que isso não é uma seita. É uma pergunta errada.

Basta analisar os processos de raciocínio do grupo por si mesmo e decidir se é algo do qual você deseja fazer parte, uma vez que tenha superado o medo da estranheza. É sua responsabilidade pessoal impedir-se de pensar de maneira sectária, não importa em qual grupo você esteja envolvido atualmente.

Quando alguém pergunta “Isso não é uma seita, é?”, não importa como eu responda, sempre sinto que estou defendendo algo. Não gosto desse sentimento. Não é função de um [mestre bayesiano](#) fornecer segurança, nem dos racionalistas defenderem.

As seitas se alimentam do pensamento de grupo, da ansiedade e do desejo de segurança. Não se

pode simplesmente fazer a ansiedade desaparecer desejando, e a falsa autoconfiança é ainda pior. Mas, enquanto alguém precisar de garantias - mesmo sendo um racionalista - isso sempre será uma falha em sua armadura. Um espadachim habilidoso foca no alvo, em vez de desviar o olhar para ver se alguém está rindo. Quando você sabe o que está tentando fazer e por quê, saberá se está tendo sucesso ou não, e se um grupo está ajudando ou atrapalhando.

(P.S.: Se alguém se aproximar de você e disser: “Tem certeza de que isso não é uma seita?” não tente explicar todos esses conceitos de uma só vez. Você está subestimando as distâncias inferenciais. Essa pessoa vai dizer: “Ahá, então você está admitindo ser uma seita!” ou “Espere, você está dizendo que eu não deveria me preocupar em entrar para uma seita?” ou “Então... o medo de seitas é uma seita para mim.” Portanto, o último fator irritante — número 7, se você está contando — é que tudo isso é uma longa história para explicar.)



## **K - Deixar ir**

## 121 - A importância de dizer “Opa!”



Acabei de terminar de ler uma história sobre a queda da Enron, *The Smartest Guys in the Room* (Os Caras Mais Espertos da Sala). Este livro ganha meu prêmio de “Título de Livro Menos Apropriado”.

Uma característica não surpreendente da lenta deterioração e abrupto colapso da Enron foi que os executivos nunca admitiram ter cometido um grande erro. Quando a catástrofe n.º 247 crescia a tal ponto que exigia uma mudança de política real, eles diziam: “Que pena que não deu certo, era uma ideia tão boa — como vamos esconder o problema em nosso balanço patrimonial?” Em vez de dizer: “Agora parece óbvio, olhando para trás, que foi um erro desde o começo” ou “Eu fui um idiota”. Nunca houve um momento de virada, um momento de reconhecimento humilde, de admitir um problema fundamental. Mesmo depois da falência, Jeff Skilling, ex-COO e breve CEO da Enron, recusou o conselho de seus próprios advogados de se valer da Quinta Emenda; ele testemunhou perante o Congresso que a Enron havia sido uma grande empresa.

Nem toda mudança é uma melhoria, mas toda melhoria é necessariamente uma mudança. Se só admitirmos pequenos erros locais, faremos apenas pequenas mudanças locais. A motivação para uma grande mudança vem do reconhecimento de um grande erro.

Na minha infância, fui criado com uma dose igual de ciência e ficção científica, e desde Heinlein até Feynman, aprendi os princípios da Racionalidade Tradicional: as teorias devem ser ousadas e se expor à falsificação; esteja disposto a cometer o sacrifício heroico de abandonar suas próprias ideias quando confrontado com evidências contrárias; seja gentil em seus argumentos; tente não se enganar; e outras ideias vagas.

Uma criação racionalista tradicional tenta produzir argumentadores que, eventualmente, cederão às evidências contrárias — deve haver alguma montanha de evidências suficientes para movê-los. Isso não é trivial; isso distingue a ciência da religião. Mas há menos foco na velocidade, em abandonar a luta o mais rápido possível, integrando as evidências eficientemente para ser necessário apenas um mínimo de evidências contrárias para destruir sua crença querida.

Fui criado na Racionalidade Tradicional e achava que era bastante racionalista. Mudei para o *Bayescraft* (Laplace/Jaynes/Tversky/Kahneman) após... bem, é uma longa história. Basicamente, mudei porque percebi que as ideias vagas da Racionalidade Tradicional não foram suficientes para me impedir de cometer um grande erro.

Depois de final e completamente admitir meu erro, olhei para trás no caminho que me levou à minha terrível descoberta. E vi que havia feito uma série de pequenas concessões, concessões mínimas, concedendo cada milímetro de terreno a contragosto, percebendo o mínimo possível do meu erro em cada ocasião, admitindo a falha apenas em pequenas doses toleráveis. Percebi que poderia ter me movido muito mais rápido se tivesse simplesmente gritado “OPA!”

E pensei: preciso elevar o nível do meu jogo.

Há uma vantagem poderosa em admitir que você cometeu um grande erro.

É doloroso. Também pode mudar toda a sua vida.

É importante ter o momento decisivo, o momento de humilde compreensão. Reconhecer um problema fundamental, não o dividir em pequenos erros palatáveis.

Não se entregue ao drama e não se orgulhe de admitir erros. Certamente, é melhor acertar na primeira vez. Mas se cometer um erro, é muito melhor reconhecê-lo de uma vez. Até mesmo do ponto de vista hedônico, é melhor sofrer uma grande perda do que muitas pequenas. A alternativa é prolongar a batalha consigo mesmo durante anos. A alternativa é a Enron.

Desde então, tenho observado outras pessoas fazendo sua própria série de concessões mínimas, cedendo a cada milímetro de terreno relutantemente; nunca confessando um erro global quando um local é suficiente; sempre aprendendo o mínimo possível com cada erro. O que poderiam consertar voluntariamente de uma só vez, transformam em pequenos remendos locais pelos quais precisam argumentar. Nunca dizem, após confessar um erro, “Eu fui um tolo”. Eles fazem o melhor para minimizar o constrangimento, dizendo “Eu estava certo na teoria”, ou “Poderia ter funcionado”, ou “Ainda quero abraçar a verdadeira essência daquilo a que estou apegado.” Defendendo seu orgulho neste momento passageiro, eles garantem que novamente cometerão o mesmo erro, e novamente precisarão defender seu orgulho.

Melhor engolir a pílula amarga inteira em um único gole terrível.

## 122 - A oferta absurda



Quando eu era muito jovem — acredito que com treze ou talvez quatorze anos —, pensei ter encontrado uma refutação para o Argumento da Diagonal de Cantor, um famoso teorema que demonstra que os números reais são mais numerosos que os números racionais. Ah, os sonhos de fama e glória que dançavam em minha mente!

Minha ideia era que, como cada número inteiro pode ser decomposto em um conjunto de potências de 2, seria possível mapear os números inteiros para o conjunto de subconjuntos de números inteiros simplesmente escrevendo a expansão binária. Por exemplo, o número 13, representado como 1101 em binário, seria mapeado para o conjunto  $\{0, 2, 3\}$ . Levei uma semana inteira para perceber que talvez eu devesse aplicar o Argumento da Diagonal de Cantor à minha construção inteligente e, é claro, ele encontrou um contraexemplo — o número binário  $(\dots 1111)$ , que não corresponde a nenhum número inteiro finito.

Foi então que encontrei esse contra exemplo e percebi que minha tentativa de refutação era falsa, com meus sonhos de fama e glória.

Inicialmente, fiquei um pouco desapontado.

Aquele pensamento passou pela minha cabeça: “Um dia, vou conseguir provar esse teorema! Mesmo que minha primeira tentativa tenha falhado, algum dia refutarei o Argumento da Diagonal de Cantor!” Ressenti-me do teorema por ser teimosamente verdadeiro, por me privar de fama e fortuna, e comecei a procurar outras refutações.

Então, percebi algo. Percebi haver cometido um erro e que, agora que havia identificado meu equívoco, não havia absolutamente nenhuma razão para duvidar da força do Argumento da Diagonal de Cantor mais do que de outros importantes teoremas da matemática.

Vi claramente que estava sendo oferecida a oportunidade de me tornar um lunático em matemática e passar o resto da vida escrevendo cartas raivosas em tinta verde para professores de matemática. (Uma vez li um livro sobre excêntricos matemáticos.)

Eu não queria que esse fosse o meu futuro, então soltei uma risadinha e deixei para lá. Aceitei o Argumento da Diagonal de Cantor com votos de felicidade e nunca mais o questioneei.

Não me lembro agora se pensei isso naquele momento ou se foi algo que ocorreu posteriormente... mas que teste terrivelmente injusto para uma criança de treze anos. O fato de eu ter que ser tão racional, tão cedo, ou falharia.

Quanto mais inteligente você for, mais jovem pode ser quando tiver a sensação de ter uma ideia verdadeiramente revolucionária pela primeira vez. Tive a sorte de ter presenciado meu próprio erro, de não precisar que outro matemático me apontasse e talvez me desse uma fonte externa para culpar. Tive sorte porque a refutação era simples o suficiente para que eu pudesse compreendê-la. Talvez eu tivesse me recuperado eventualmente, caso contrário. Já me recuperei de coisas piores quando adulto. Mas, se tivesse errado tão cedo, teria desenvolvido essa habilidade?

Fico pensando em quantas pessoas, quando tinham treze anos e cometeram seu primeiro erro fatal, começaram a escrever cartas raivosas com tinta verde. Quantas mentes promissoras foram destruídas antes disso?

Eu cometi um erro. Apenas isso. Eu não estava realmente certo, no fundo; não foi uma vitória moral; não estava demonstrando ambição, ceticismo ou qualquer outra virtude maravilhosa; não foi um erro razoável; eu não estava parcialmente certo, nem mesmo levemente certo. Tive um pensamento que nunca teria tido se fosse mais sábio, e isso é tudo.

Se não tivesse conseguido admitir isso para mim mesmo, se tivesse reinterpretado meu erro como algo virtuoso, se tivesse insistido em estar pelo menos um pouco certo por orgulho, então não teria desistido. Teria continuado procurando uma falha no Argumento da Diagonal. E, mais cedo ou mais tarde, talvez tivesse encontrado.

Enquanto não [admitirmos estarmos errados](#), não podemos prosseguir com nossas vidas; nossa auto-imagem ainda estará ligada ao antigo erro.

Sempre que somos tentados a nos apegar a um pensamento que nunca teríamos tido se fôssemos mais sábios, é-nos oferecida a oportunidade de nos tornarmos insanos — mesmo que nunca escrevamos cartas raivosas com tinta verde. Mesmo que ninguém discuta com você, mesmo que nunca conte a ninguém sobre sua ideia, você ainda pode estar insano. É o apego que o define.

Não é verdade. No fundo, não é verdade. Não é uma meia-verdade nem uma fração de verdade. É apenas um pensamento que você nunca deveria ter tido. Nem todo lado tem um aspecto positivo. Os seres humanos cometem erros, e nem todos são fracassos disfarçados de sucesso. Os seres humanos cometem erros; isso acontece, só isso. Diga “opa!” e siga em frente com sua vida.



## 123 - Simplesmente perca as esperanças de uma vez



Casey Serin, um programador de 24 anos sem experiência prévia no setor imobiliário, [deve 2,2 milhões de dólares](#) aos bancos após ter mentido em pedidos de hipoteca para comprar simultaneamente oito casas diferentes em estados diferentes. Ele obteve dinheiro através das hipotecas, solicitando quantias superiores aos valores das casas, e gastou esse dinheiro em despesas com moradia e seminários imobiliários. Aparentemente ele tinha a esperança de que o mercado imobiliário valorizasse.

No entanto, a parte mais triste não é essa. A parte mais triste é que ele não desistiu. Casey Serin não aceita a derrota. Ele se recusa a declarar falência ou conseguir um emprego; ele [ainda acredita](#) que pode ter sucesso no setor imobiliário. Ele continua a gastar dinheiro em seminários e até tentou hipotecar uma nona casa. Ele não falhou, entenda, ele apenas teve uma experiência de aprendizado.

Isso é o que acontece quando se recusa a perder a esperança.

Embora esse comportamento possa parecer simplesmente estúpido, também me faz lembrar de dois economistas ganhadores do Prêmio Nobel...

... ou seja, Merton e Scholes da [Long-Term Capital Management \(LTCM\)](#).

Apesar de a LTCM ter obtido lucros enormes nos primeiros três anos, em 1998 as ineficiências que a LTCM explorava começaram a desaparecer — outras pessoas conheciam o truque, então ele deixou de funcionar.

A LTCM recusou-se a perder a esperança. Viciados em retornos anuais de 40%, eles emprestaram cada vez mais, alavancando-se para explorar margens decrescentes. Quando tudo começou a dar errado para a LTCM, eles tinham um patrimônio líquido de US\$4,72 bilhões, uma alavancagem de US\$124,5 bilhões e posições em derivativos de US\$1,25 trilhão.

Cada profissão tem sua própria forma de inteligência — diferentes habilidades a aprender e regras a seguir. Assim, você pode pensar que o estudo da “racionalidade” como disciplina geral não teria muito a contribuir para o sucesso na vida real. No entanto, parece-me que evitar ser estúpido tem muito em comum em todas as profissões. Se você pretende ensinar alguém a não transformar pequenos erros em grandes erros, é quase a mesma arte, seja em fundos de *hedge* ou na escrita de romances. E uma das chaves é esta: esteja pronto para admitir a derrota.

## 124 - O uso adequado da dúvida



Certa vez, durante uma discussão sobre o Caminho, notei como a maioria dos sistemas de crenças organizados tende a evitar a dúvida. Um ouvinte respondeu que os jesuítas deveriam ser imunes a essa crítica, ao praticarem a dúvida organizada: os noviços são instruídos a duvidar do cristianismo, da existência de Deus, da validade de seu chamado e até mesmo de sua capacidade de cumprir os votos perpétuos de castidade e pobreza. Então perguntei: “Bem, eles duvidam para superar essas dúvidas, certo?” E ele respondeu: “Não, eles devem duvidar que suas próprias dúvidas possam crescer e se fortalecer”.

Não consegui confirmar ou refutar essas alegações (se alguém puder ajudar, ficarei muito grato). No entanto, acho esse cenário fascinante e digno de discussão, independentemente de sua veracidade em relação aos jesuítas. Se eles praticassem essa dúvida deliberada, conforme descrito, seriam virtuosos como racionalistas?

Admito que, no cenário (possivelmente hipotético) acima, os jesuítas não seriam adequadamente descritos como “fugindo da dúvida”. Contudo, essa conduta (possivelmente hipotética) ainda me parece altamente suspeita. Para um racionalista verdadeiramente virtuoso, a dúvida não deveria ser assustadora. A conduta descrita parece um programa de dessensibilização para algo extremamente assustador, como expor um aracnofóbico à aranhas sob condições cuidadosamente controladas.

Mas, mesmo assim, eles estão encorajando seus noviços a duvidar, certo? Importa se suas razões são falhas? Isso não seria uma ação digna de um racionalista?

Toda curiosidade busca sua própria aniquilação; não há curiosidade que não deseje uma resposta. Mas quando você obtém uma resposta e satisfaz sua curiosidade, o glorioso mistério deixa de ser misterioso.

Da mesma forma, toda dúvida existe para aniquilar uma crença específica. Se uma dúvida não consegue destruir seu alvo, ela não foi satisfeita — mas ainda é uma resolução, um fim, embora mais triste. Uma dúvida que não se destrói nem destrói seu alvo provavelmente nunca existiu. É a resolução das dúvidas, não apenas o mero ato de duvidar, que impulsiona a catraca da racionalidade.

Toda melhoria é uma mudança, mas nem toda mudança é uma melhoria. Todo racionalista tem dúvidas, mas nem toda dúvida é racional. Vestir-se de dúvidas não faz de você um racionalista, assim como vestir um jaleco branco não faz de você um médico.

Uma dúvida racional surge por uma razão específica — você tem justificativa para suspeitar que a crença está errada. Essa razão implica uma investigação que irá destruir a crença ou destruir a dúvida. Isso também se aplica a dúvidas abstratas, como “Será que há uma hipótese mais simples que explique esses dados?”. Nesse caso, você investiga pensando em hipóteses mais simples. À medida que a busca por respostas não traz resultados, você percebe que é cada vez menos provável que a próxima tentativa seja bem-sucedida. Eventualmente, o custo da pesquisa supera o benefício esperado e você para de procurar. Nesse ponto, não se pode mais afirmar que está duvidando de maneira útil. Uma dúvida não investigada é como se não existisse. Toda dúvida existe para ser destruída, de uma maneira ou de outra. Uma dúvida não resolvida é uma ação nula; não faz a roda girar, nem para frente, nem para trás.

Se você realmente acredita em uma religião (não apenas acredita nela), por que pediria aos noviços que considerem dúvidas que não serão satisfeitas? Seria como pedir aos estudantes de física que duvidem meticulosamente que a revolução do século XX possa ter sido um erro, e que a mecânica newtoniana esteve

correta o tempo todo. Se você realmente não duvida de algo, por que fingir que duvida?

Porque todos queremos ser vistos como racionais, e a dúvida é amplamente considerada uma virtude do racionalista. No entanto, não se entende amplamente que você precisa de um motivo específico para duvidar e que uma dúvida não resolvida é uma ação nula. Em vez disso, as pessoas pensam que isso é modéstia, um comportamento submisso, mantendo a hierarquia do status tribal — é quase o mesmo problema da [humildade, sobre o qual escrevi anteriormente](#). Fazer uma exibição pública de dúvida para se convencer de que você é racional é tão vazio quanto usar um jaleco.

Para evitar professar dúvidas, lembre-se:

- Uma dúvida racional existe para destruir a crença alvo e, se não a destruir, ela morre insatisfeita.
- Uma dúvida racional surge por alguma razão específica pela qual a crença pode estar errada.
- Uma dúvida não resolvida é uma ação nula.
- Uma dúvida não investigada pode muito bem não existir.
- Não se deve se orgulhar de meras dúvidas, mas pode-se orgulhar justamente ao despedaçar uma crença acalentada.
- Embora possa exigir coragem para enfrentar suas dúvidas, nunca se esqueça de que, para uma mente ideal, a dúvida não seria assustadora para começo de conversa.

## 125 - Você pode encarar a realidade



O que é verdadeiro já é assim.

Reconhecê-lo não o torna pior.

Não o admitir não o faz desaparecer.

E, porque é verdadeiro, é o que está lá para interagirmos.

Nada que seja falso está lá para ser vivido.

As pessoas conseguem suportar o que é verdadeiro, pois já o estão suportando<sup>57</sup>.

— Eugene Gendlin

---

57 NT.: Texto original em inglês. *What is true is already so. Owning up to it doesn't make it worse. Not being open about it doesn't make it go away. And because it's true, it is what is there to be interacted with. Anything untrue isn't there to be lived. People can stand what is true, for they are already enduring it.*

## 126 - A Meditação sobre a Curiosidade



A primeira virtude é a curiosidade.

— “As Doze Virtudes da Racionalidade”

Como racionalistas, somos obrigados a criticar e questionar nossas crenças... certo?

Considere o que acontece com você, em um nível psicológico, se começar dizendo: “É meu dever criticar minhas próprias crenças”. Roger Zelazny já diferenciou entre “querer ser um autor” e “querer escrever”. Mark Twain disse: “Um clássico é algo que todo mundo quer ter lido e ninguém quer ler”. Criticar a si mesmo por senso de dever faz você desejar ter investigado, para poder dizer depois que sua fé não é cega. Isso não é o mesmo que querer investigar.

Isso pode levar a uma [interrupção motivada](#) da sua investigação. Você considera uma objeção, depois um contra-argumento para essa objeção, e então para por aí. Você repete isso com várias objeções, até sentir que cumpriu sua obrigação de investigar, e para por aí. Você alcançou seu objetivo psicológico subjacente: livrar-se da dissonância cognitiva que resultaria de se pensar como um racionalista, mas saber que não tentou criticar sua crença. Você poderia chamar isso de compra de satisfação racionalista—tentando criar uma “sensação agradável” de dever cumprido.

Depois, seu nível de probabilidade declarado será alto o suficiente para justificar a manutenção dos planos e crenças com os quais você começou, mas não tão alto a ponto de evocar incredulidade de si mesmo ou de outros racionalistas.

Quando você está realmente curioso, tenderá a investigar as perguntas que parecem mais promissoras para produzir mudanças em sua crença, ou as perguntas menos parecidas com as que você já tentou antes. Depois, a distribuição da sua probabilidade provavelmente não se parecerá com a que você tinha quando começou - mudanças devem ter ocorrido, seja para cima ou para baixo, e qualquer direção é igualmente aceitável se você estiver realmente curioso. Contraste isso com o motivo subconsciente de manter sua investigação em terreno familiar, para que você possa terminá-la rapidamente, para poder ter investigado e restaurar o equilíbrio familiar em que seus velhos planos e crenças se baseiam.

Quanto ao que eu acho que a verdadeira curiosidade deve parecer e o poder que ela possui, eu o remeto a “Uma Fábula da Ciência e da Política”. Cada um dos personagens destina-se a ilustrar diferentes lições. Ferris, o último personagem, personifica o poder da curiosidade inocente: a leveza e um desejo ansioso por evidências.

Ursula K. Le Guin escreveu:

“Na inocência, não há força contra o mal. Mas há força nela para o bem<sup>58</sup> [1].

A curiosidade inocente pode se desviar inocentemente; e, portanto, o treinamento de um racionalista, e sua [sofisticação](#) acompanhante, deve ser desafiado como um perigo se quisermos nos fortalecer. No entanto, podemos tentar manter a leveza e o desejo ansioso da inocência.

---

58 NT.: Texto original em inglês. *In innocence there is no strength against evil. But there is strength in it for good*

Como está escrito em “As Doze Virtudes da Racionalidade”:

Se em seu coração você acredita que já sabe, ou se em seu coração você não deseja saber, então seu questionamento será sem propósito e suas habilidades sem direção. A curiosidade busca se aniquilar; não há curiosidade que não queira uma resposta.

Não há substituto para a curiosidade genuína. Um desejo ardente de saber é mais forte do que uma solene promessa de buscar a verdade. Mas você não pode produzir curiosidade apenas querendo, assim como não pode fazer seu pé sentir calor quando está frio. Às vezes, tudo o que temos são nossas meras promessas solenes.

Então, o que podemos fazer com o dever? Para começar, podemos tentar ter interesse em nossas investigações por dever — manter um olhar atento por faíscas de intrigas genuínas ou até mesmo uma ignorância genuína e um desejo de resolvê-la. Isso se alinha perfeitamente com manter um olhar especial para as possibilidades [dolorosas](#), pelas quais você está recuando — nem tudo é pensamento negativo.

Meditar sobre a “Conservação de Evidência Esperada” também deve ajudar. Para cada novo ponto de investigação, para cada peça de evidência não vista que você de repente observa, a probabilidade posterior esperada deve ser igual à probabilidade anterior. No microprocesso da investigação, sua crença deve sempre estar equilibrada para mudar em qualquer direção. Nem todo ponto pode ser suficiente para abrir amplamente a questão, mudando a crença de 70% para 30% de probabilidade, mas se sua crença atual é de 70%, você deve estar tão pronto para diminuí-la para 69% quanto para aumentá-la para 71%. Você não deve pensar que sabe em qual direção irá (em média), porque pelas leis da teoria da probabilidade, se você sabe seu destino, já chegou lá. Se você puder investigar honestamente, de forma que cada novo ponto tenha realmente o potencial igual de mudar a crença para cima ou para baixo, isso pode auxiliar a mantê-lo interessado ou até mesmo curioso sobre o microprocesso da investigação.

Se o argumento que você está considerando não é novo, por que sua atenção está indo para lá? É lá que você olharia se fosse realmente curioso? Você está criticando subconscientemente sua crença em seus pontos fortes, em vez de seus pontos fracos? Você está [revisando a evidência](#)?

Se você conseguir *não* revisar o suporte já conhecido e conseguir reduzir sua crença por uma pequena porção de cada vez com base em novas evidências, poderá até mesmo conseguir abandonar completamente a crença — percebendo de qual direção os ventos da evidência estão soprando contra você.

Outra forma de restaurar a curiosidade é o que eu chamo de Litania de Tarski, a qual é realmente uma meta-litania que se especializa para cada instância (isso é apropriado). Por exemplo, se eu estiver ansioso imaginando se uma caixa trancada contém um diamante, ao invés de pensar em todas as possíveis consequências maravilhosas se a caixa realmente contiver um diamante, eu posso repetir a Litania de Tarski:

*Se a caixa contém um diamante,*

*Desejo acreditar que a caixa contém um diamante;*

*Se a caixa não contém um diamante,*

*Desejo acreditar que a caixa não contém um diamante; Que eu não me apegue a crenças que talvez eu não queira.*

Então, você deveria meditar sobre a possibilidade de não haver um diamante, e a vantagem subsequente que virá se você acreditar não haver diamante, e a desvantagem subsequente se você acreditar que há um diamante. Veja também a [Litania de Gendlin](#).

Se você puder encontrar dentro de si o menor vestígio de verdadeira incerteza, então guarde-o como um guarda-florestal cuidando de uma fogueira. Se você conseguir fazer essa incerteza se inflamar em uma chama de curiosidade, isso o tornará ávido e ansioso, e dará propósito às suas perguntas e direção às suas habilidades.

## Referências

[1] Ursula K. Le Guin, *The Farthest Shore* (Saga Press, 2001).

## 127 - Ninguém pode isentá-lo das leis da racionalidade



A Racionalidade Tradicional é formulada em termos de regras sociais, nas quais violações são interpretadas como trapaça, como quebras das normas cooperativas. Se você quer que eu aceite uma crença sua, é obrigado a me fornecer uma quantidade adequada de evidências. Se você tentar escapar dessa obrigação, todos sabemos que está trapaceando e descumprindo sua responsabilidade. Uma teoria deve fazer previsões arrojadas por si mesma, não apenas roubar previsões feitas por outras teorias. Uma teoria precisa se expor à falsificação — se ela tentar se esquivar, é como tentar evitar um temível ritual de iniciação; você deve pagar suas dívidas.

A Racionalidade Tradicional é formulada de maneira semelhante aos costumes que regem as sociedades humanas, facilitando sua transmissão oral. Os seres humanos detectam trapaças sociais com muito mais confiabilidade do que violações isomórficas de regras lógicas abstratas. No entanto, considerar a racionalidade como uma obrigação social leva a algumas ideias estranhas.

Por exemplo, encontramos pessoas religiosas defendendo suas crenças dizendo: “Bem, você não pode justificar a sua crença na ciência!” Em outras palavras, “Como você se atreve a me criticar por ter crenças não justificadas, seu hipócrita! Você também está fazendo isso!”

Para os bayesianos, o cérebro é um mecanismo de precisão: ele processa e concentra evidências emaranhadas em um mapa que reflete o território. Os princípios da racionalidade são leis no mesmo sentido que a Segunda Lei da Termodinâmica: para obter uma crença confiável, é necessária uma quantidade calculável de evidências emaranhadas, assim como para resfriar confiavelmente o conteúdo de uma geladeira é necessário um mínimo calculável de energia livre.

Em princípio, as leis da física são reversíveis no tempo, então há uma probabilidade infinitesimalmente pequena — indistinguível de zero para todos, exceto para os matemáticos — de uma geladeira esfriar espontaneamente enquanto gera eletricidade. Há uma chance infinitesimalmente maior de você conseguir desenhar com precisão um mapa detalhado das ruas de Nova York sem nunca a visitar, sentado em sua sala de estar com as cortinas fechadas e sem conexão com a internet. Mas eu não ficaria esperando por isso.

Antes de tentar mapear um território invisível, coloque um pouco de água em um copo à temperatura ambiente e espere até que ela congele espontaneamente antes de prosseguir. Dessa forma, você pode ter certeza de que o truque geral — ignorar probabilidades infinitesimalmente pequenas de sucesso — está funcionando corretamente. Você pode não perceber diretamente que o seu mapa está errado, especialmente se nunca visitou Nova York, mas você pode ver que a água não congela sozinha.

Se as regras da racionalidade são costumes sociais, pode parecer justificar o comportamento X se apontarmos que outros estão fazendo a mesma coisa. Não seria justo exigir provas de você se nós mesmos não pudermos fornecê-las. Reconheceríamos que nenhum de nós é melhor do que os outros, e iríamos ceder e misericordiosamente desobrigá-lo da sua responsabilidade social de fornecer evidências para a sua crença. E todos viveríamos felizes para sempre em liberdade, fraternidade e igualdade.

Se as regras da racionalidade são leis matemáticas, tentar justificar uma crença sem evidências, apontando para outra pessoa que está fazendo o mesmo, será tão eficaz quanto listar trinta razões pelas quais você não deve cair de um penhasco.

Mesmo que todos votemos que é injusto que uma geladeira precise de eletricidade, ela ainda não



funcionará (com probabilidade  $\sim 1$ ). Mesmo que todos votemos que você não deveria visitar Nova York, o mapa continuará errado. A Natureza é notoriamente indiferente a tais súplicas, assim como a Matemática.

Portanto, retornando à linguagem social da Racionalidade Tradicional, não pense que pode escapar afirmando que é aceitável ter crenças arbitrárias sobre XYZ porque outras pessoas também as têm. Se duas partes de um contrato se comportarem igualmente mal, um juiz humano pode decidir não impor penalidades a nenhuma delas. Mas se dois engenheiros projetarem seus motores igualmente mal, nenhum deles funcionará. Um erro de projeto não pode desculpar outro. Mesmo que eu esteja fazendo XYZ errado, isso não o ajuda ou o isenta das regras; significa apenas que estamos em maus lençóis.

Como uma questão de lei humana em democracias liberais, todos têm direito às suas próprias crenças. Por uma questão de lei natural, você não tem direito à precisão. Não punimos pessoas por acreditarem em coisas estranhas, pelo menos não nos países mais sábios. Mas ninguém pode revogar a lei de que você precisa de evidências para formar crenças precisas. Nem mesmo um voto de toda a espécie humana pode obter misericórdia no tribunal da Natureza.

Físicos não decidem as leis da física, apenas especulam sobre quais são. Racionalistas não decidem as leis da racionalidade, apenas especulamos sobre quais são. Você não pode “[racionalizar](#)” qualquer coisa que não seja racional desde o início. Se, por meio de uma persuasão extraordinária, você convencer todos os físicos do mundo de que está isento da lei da gravidade e pular de um penhasco, você cairá. Até mesmo dizer “nós não decidimos” é muito antropomórfico. Não há autoridade superior que possa isentá-lo. Apenas há causa e efeito.

Lembre-se disso quando implorar para ser dispensado apenas desta vez. Não podemos desculpá-lo. Isso não está em nossas mãos.

## 128 - Deixe uma linha de retirada



“Ao cercar o inimigo,  
Sempre permita uma rota de fuga.  
Eles devem ver que há  
Uma alternativa à morte<sup>59</sup>.”

— Sun Tzu, *A Arte da Guerra* [1]

“Não aumente a pressão, abaixe a parede<sup>60</sup>.”

— Lois McMaster Bujold, *Komarr* [2]

Certa vez, conversei com alguém que não era racionalista e que, por acaso, havia entrado em uma reunião local de racionalistas. Essa pessoa havia acabado de expressar (a) sua crença em almas e (b) sua descrença em criogenia, pois acreditava que a alma não permaneceria no corpo congelado. Eu perguntei: “Mas como você sabe disso?” Pela confusão que surgiu em seu rosto, ficou claro que essa pergunta nunca havia passado por sua mente. Não digo isso pejorativamente — ela parecia uma pessoa legal, mas sem absolutamente nenhum treinamento em racionalidade, assim como a maioria das pessoas. Eu realmente preciso escrever esse livro.

A maior parte da conversa que se seguiu foi sobre tópicos já abordados no *Overcoming Bias* — se você estiver realmente curioso sobre algo, provavelmente poderá encontrar uma maneira adequada de testá-lo; tente obter crenças precisas primeiro e deixe que suas emoções sejam influenciadas por isso — coisas desse tipo. No entanto, a conversa me fez lembrar de uma ideia que ainda não abordei aqui:

“Sugiro que você tente visualizar como seria o mundo se não houvesse almas e o que você faria a respeito disso,” eu disse a ela. “Não pense em todas as razões pelas quais isso não poderia ser verdade, apenas aceite como uma premissa e visualize as consequências. Pense em coisas como: ‘Bem, se não houver almas, posso me inscrever para a criogenia’ ou ‘Se não houver Deus, posso continuar sendo moral mesmo assim’, em vez de enfrentar uma realidade muito assustadora. Por uma questão de autoestima, você deve tentar acreditar, na verdade, mesmo que seja desconfortável, como eu disse antes. No entanto, por uma questão de natureza humana, ajuda tornar uma crença menos desconfortável antes de tentar avaliar as evidências a favor dela.”

O princípio por trás dessa técnica é simples: assim como Sun Tzu aconselha a fazer com seus inimigos, você também deve fazer consigo mesmo — deixar uma rota de fuga para que você tenha menos dificuldade em recuar. A perspectiva de perder o emprego, por exemplo, pode parecer muito mais assustadora quando você sequer consegue pensar nisso, do que após calcular exatamente quanto tempo suas economias durarão, verificar o mercado de trabalho em sua área e planejar o que fazer em seguida. Somente assim você estará preparado para avaliar adequadamente a probabilidade de manter seu emprego durante os próximos cortes planejados. Seja um verdadeiro covarde e planeje sua saída em detalhes — visualize cada passo — de

---

59 NT.: Texto original em inglês. *When you surround the enemy; always allow them an escape route. They must see that there is; an alternative to death.*

60 NT.: Texto original em inglês. *Don't raise the pressure, lower the wall.*

preferência antes mesmo de entrar no campo de batalha.

A esperança é que seja menos difícil ter coragem de visualizar um cenário desconfortável como um experimento mental do que considerar a probabilidade de ser verdade. No entanto, após fazer o primeiro, fica mais fácil fazer o segundo.

Lembre-se de que o bayesianismo é preciso — mesmo que uma proposição assustadora pareça improvável, ainda é importante considerar todas as evidências, a favor e contra, de maneira justa, para chegar a uma probabilidade quantitativa racional. Visualizar uma crença assustadora não significa admitir que você acredita, no fundo, que ela é provavelmente verdadeira. Você pode visualizar uma crença assustadora com base nos princípios gerais de uma boa gestão mental. “O pensamento que você não consegue pensar controla mais você, do que os pensamentos que você expressa em voz alta” — isso acontece mesmo quando o pensamento impensável é falso!

A técnica de deixar uma linha de retirada requer um mínimo de auto-honestidade para ser utilizada corretamente.

Primeiramente, é preciso poder admitir para si mesmo quais ideias o assustam e a quais ideias você está apegado. No entanto, esse é um teste substancialmente menos difícil do que contar com justiça as evidências de uma ideia que o assusta. Seria útil se eu dissesse que já tive a oportunidade de usar essa técnica eu mesmo? Afinal, um racionalista não rejeita todas as emoções. Existem ideias que me assustam, mas ainda acredito que sejam falsas. Existem ideias às quais sei que estou apegado, mas ainda acredito que sejam verdadeiras. No entanto, ainda assim, planejo meus “retiros”, não porque estou planejando me retirar, mas porque planejar meu “retiro” com antecedência me ajuda a pensar sobre o problema sem apego.

Mas o maior teste de auto-honestidade é realmente aceitar a proposição desconfortável como uma premissa e descobrir como você lidaria realmente com isso. Quando nos deparamos com uma ideia desconfortável, nosso primeiro impulso é naturalmente pensar em todas as razões pelas quais isso não pode ser verdade. Assim, você encontrará uma certa quantidade de resistência psicológica em si, se tentar visualizar exatamente como o mundo seria e o que você faria a respeito se a sua crença mais estimada fosse falsa ou se a sua crença mais temida fosse verdadeira.

Pense em todas as pessoas que afirmam que, sem Deus, a moralidade seria impossível. (E sim, esse tópico surgiu na conversa, então não estou criando um espantalho.) Se os teístas pudessem visualizar sua reação real ao acreditar como um fato que Deus não existe, eles poderiam perceber que não, eles não sairiam por aí matando bebês. Eles poderiam perceber que os ateus estão reagindo à inexistência de Deus da mesma forma que eles próprios reagiriam se acreditassem nisso. Eu menciono isso para ilustrar que é um desafio considerável visualizar como você realmente reagiria acreditando no oposto de uma crença fortemente arraigada. Além disso, é sempre contra-intuitivo perceber que, sim, as pessoas superam coisas.

As pessoas recém-tetraplégicas não ficam tão tristes depois de seis meses como esperavam, por exemplo. Tetraplégicos lidam com a situação, e você também lidaria.

Considere também a [Litania de Gendlin](#) e a [Litania de Tarski](#). O que é verdade já é verdade; admitir isso não piora as coisas. Não deve haver medo em apenas visualizar um mundo que você teme. Se esse mundo já é real, visualizá-lo não o tornará pior; e se não é real, visualizá-lo não fará mal. E lembre-se, ao visualizar, que se as coisas assustadoras que você está imaginando forem realmente verdadeiras — podendo não ser o caso! — então você gostaria realmente de acreditar nisso, e você deveria visualizar isso também; não acreditar nisso não o ajudaria.

Quantas pessoas religiosas manteriam sua crença em Deus se pudessem visualizar com precisão aquele mundo hipotético em que Deus não existe e elas próprias se tornaram ateias?

Deixar uma linha de retirada é uma técnica poderosa, mas não é fácil. A visualização honesta não exige tanto esforço quanto admitir abertamente que Deus não existe, mas ainda assim requer esforço.

## Referências

[1] Sun Tzu, The Art of War (Cloud Hands, Inc., 2004).

[2] Lois McMaster Bujold, Komarr, Miles Vorkosigan Adventures (Baen, 1999).

## 129 - Crise de fé

“Não se pode considerar uma verdadeira crise de fé a menos que as coisas possam pender facilmente para qualquer lado<sup>61</sup>.”

— [Thor Shenkel](#)

Muitas pessoas mantêm crenças cujas falhas até uma criança de dez anos poderia apontar, se as ouvisse pela primeira vez. Não estamos falando de erros sutis. Esses equívocos seriam óbvios para uma mente [imparcial](#) que [não evitasse](#) o ceticismo infantil. Como disse o Verificador de Premissas, “se a ideia de Deus não tivesse surgido até a era científica, apenas uma pessoa excepcionalmente estranha inventaria tal ideia e fingiria que ela explicava alguma coisa”.

No entanto, mesmo cientistas especialistas, incluindo os principais inovadores em seus campos, não aplicam esse ceticismo com sucesso. O ganhador do Prêmio Nobel, Robert Aumann, famoso pelo Teorema do Acordo de Aumann, é um judeu ortodoxo. Arrisco dizer que Aumann, em algum momento, deve ter questionado sua fé. No entanto, ele não conseguiu duvidar [com sucesso](#). [Mudamos de ideia com menos frequência do que pensamos](#).

Isso deveria nos assustar até a medula. Significa que mesmo sendo um cientista de renome mundial, familiarizado com a matemática bayesiana, você pode falhar em rejeitar uma crença cujo absurdo seria óbvio para uma criança de dez anos. Isso revela a defesa invencível que uma crença pode construir para si mesma quando infecta a mente há muito tempo.

O que é preciso para derrubar um erro que construiu uma fortaleza em torno de si?

No entanto, quando você sabe que é um erro, ele já está derrotado. O dilema não é ‘Como posso rejeitar a crença falsa de longa data X?’, mas sim ‘Como posso saber se a crença de longa data X é falsa?’ A auto-honestidade é mais frágil quando não temos certeza sobre qual caminho é o correto. E assim surge a pergunta:

Como podemos criar em nós mesmos uma verdadeira crise de fé, que poderia facilmente ir para qualquer lado?

A religião é o caso de julgamento que todos podemos imaginar. (Os leitores que nasceram de pais ateus perderam uma experiência fundamental na vida e devem se contentar com a pálida substituição de ponderar sobre seus amigos religiosos.) Mas se você eliminou toda simpatia e agora vê os teístas como [mutantes malignos](#), então você perdeu a capacidade de imaginar os verdadeiros conflitos internos que eles enfrentam. Você não conseguirá fazer a pergunta: “Que estratégia geral uma pessoa religiosa teria que seguir para escapar de sua religião?”

Tenho certeza de que alguns, ao enfrentar esse desafio, estão prontos para recitar uma lista de argumentos ateístas comuns — “Eles teriam que admitir que não há evidência bayesiana para a existência de Deus”, “Eles teriam que reconhecer as evasões morais que usam para desculpar o comportamento de Deus na Bíblia”, “Eles precisam aprender a aplicar a Navalha de Ocam...”

---

61 NT.: Texto original em inglês. *It ain't a true crisis of faith unless things could just as easily go either way.*

ERRADO! ERRADO! ERRADO! ERRADO!

Esse tipo de [revisão](#), em que você apenas menciona pontos nos quais já pensou bastante, é exatamente o estilo de pensamento que mantém as pessoas presas em suas religiões atuais. Se você mantiver seus [pensamentos armazenados em cache](#), se seu cérebro preencher rapidamente a resposta óbvia sem que você a [perceba originalmente](#), certamente não conseguirá passar por uma crise de fé.

Talvez seja apenas uma questão de não haver pessoas o suficiente lendo “Gödel, Escher, Bach” em uma idade suficientemente jovem, mas observei que uma grande parcela da população — inclusive pessoas tecnicamente habilidosas — tem [dificuldade em acompanhar](#) argumentos que seguem esse objetivo. Nos meus dias mais pessimistas, questiono se o camelo tem duas corcovas.

Mesmo quando é explicitamente apontado, algumas pessoas aparentemente não conseguem fazer a transição do nível do objeto “Use a navalha de Ocam! Você precisa perceber que sua crença em Deus é desnecessária!” para o meta-nível “Tente impedir que sua mente complete o padrão da maneira usual!” Porque assim como todos os seus amigos racionalistas falam sobre a Navalha de Ocam como se fosse algo bom, e assim como a Navalha de Ocam salta diretamente para sua mente também, a resposta religiosa óbvia aprovada pelos amigos é “Os caminhos de Deus são misteriosos e é presunçoso supor que podemos compreendê-los”. Então, pensar que a estratégia geral a seguir é “Use a navalha de Ocam” seria como um teísta dizendo que a estratégia geral é ter fé.

“Mas... mas a Navalha de Ocam é realmente melhor do que a fé! Não é como preferir um sabor de sorvete diferente! Qualquer pessoa pode ver, olhando para a história, que o raciocínio ocamiano tem sido muito mais produtivo do que a fé...”

Tudo isso é verdade. Mas vamos além disso. A questão é que, ao dizer isso, você está apenas repetindo uma justificativa comum que já está em sua mente. O desafio de passar por uma crise de fé é lidar com o caso em que nossas conclusões padrão e nossas justificativas usuais podem estar erradas. Portanto, se a justificativa padrão para X é “Navalha de Ocam!”, e você deseja manter uma crise de fé em relação a X, você deve questionar se a Navalha de Ocam realmente apoia X, se sua compreensão da Navalha de Ocam está correta e — se você quiser ter dúvidas realmente profundas — se a simplicidade é o critério que funcionou historicamente nesse caso, ou se seria razoável esperar que funcionasse, etc. Se você aconselharia um religioso a questionar sua crença de que “fé” é uma boa justificativa para X, então você também deve se esforçar igualmente para questionar sua crença de que “navalha de Ocam” é uma boa justificativa para X.

(Pense em todas as pessoas que não entendem o Comprimento Mínimo de Descrição ou as formulações de Indução de Solomonoff da Navalha de Ocam, que pensam que a Navalha de Ocam proíbe a existência de muitos mundos ou a [Hipótese de Simulação](#). Elas precisariam questionar suas formulações da Navalha de Ocam e suas noções sobre por que a simplicidade é algo bom. Seja qual for o X em disputa que você acabou de justificar dizendo “Navalha de Ocam!”, Aposto que não é o mesmo nível de ocorrência ocamiana que a gravidade.)

Se “Navalha de Ocam!” é a sua resposta habitual, a resposta padrão que todos os seus amigos dão, então é melhor impedir que seu cérebro complete instantaneamente esse padrão se você está tentando provocar uma verdadeira crise de fé.

É melhor seguir regras como: “Imagine o que um cético diria - e então imagine o que eles diriam em resposta à sua resposta - e então imagine o que mais eles poderiam dizer, o que seria mais difícil de responder”.

Ou “Tente pensar no pensamento que mais incomoda”. E acima de tudo, siga a regra:

“Faça o mesmo esforço desesperado que seria necessário para um teísta rejeitar sua religião”.

Porque, se você não está se esforçando tanto, então - pelo que você sabe - sua mente pode estar cheia de absurdos tão ridículos quanto a religião.

Sem um esforço agonizante e doloroso para ser racional, o tipo de esforço necessário para abandonar uma religião, como você ousa acreditar em qualquer coisa quando Robert Aumann acredita em Deus?

Alguém (não sei quem) uma vez observou que as pessoas só têm até uma certa idade para abandonar sua fé religiosa. Depois, elas terão respostas para todas as objeções, e será tarde demais. Esse é o tipo de existência que você precisa superar. Esse é um teste de sua força como racionalista, e é extremamente rigoroso; mas se você não conseguir passar, será mais fraco do que uma criança de dez anos.

Mas, novamente, quando você sabe que uma crença é um erro, ela já foi derrotada. Portanto, não estamos falando de um esforço desesperado e convulsivo para [desfazer os efeitos](#) de uma educação religiosa após chegar à conclusão de que sua religião está errada. Estamos falando de um esforço desesperado para descobrir se você deve se livrar das correntes ou mantê-las. A auto-honestidade é mais frágil quando não sabemos qual caminho seguir — é quando as racionalizações não são pecados óbvios.

- Nem toda dúvida exige uma crise de fé total. Mas você deve considerar isso quando:
- Uma crença permanece em sua mente por muito tempo;
- Você está cercado por uma nuvem de argumentos e refutações conhecidos;
- Você já [investiu demais](#) na crença (tempo, dinheiro, declarações públicas);
- A crença tem consequências emocionais (observe que isso não a torna errada);

Isso se misturou em sua personalidade em geral.

Nenhum desses sinais de alerta são refutações imediatas. Esses atributos colocam uma crença [em risco](#) de todos os tipos de perigos e tornam muito difícil rejeitá-la quando está errada. Mas também valem para a crença de Richard Dawkins na biologia evolutiva, bem como para o catolicismo do papa. Isso não quer dizer que estamos falando apenas de diferentes sabores de sorvete. Apenas os não iluminados pensam que todas as crenças profundamente arraigadas estão no mesmo nível, independentemente da evidência que as apoie, apenas porque são profundamente arraigadas. A questão não é ter crenças superficiais, mas ter um mapa que reflita o território.

Enfatizo isso, é claro, para que você possa admitir para si mesmo: “Minha crença tem esses sinais de alerta”, sem ter que dizer para si mesmo: “Minha crença é falsa”.

Mas o que esses sinais de alerta marcam é uma crença que exigirá mais do que um esforço comum para duvidar com eficácia. De modo que, se fosse de fato falso, você a rejeitaria. E onde você não pode duvidar efetivamente, você está cego, porque seu cérebro manterá a crença [incondicionalmente](#). Quando uma retina envia o mesmo sinal independentemente dos fótons que entram nela, chamamos isso de cegueira.

Quando você deve encenar uma Crise de Fé?

Mais uma vez, pense no conselho que você daria a um teísta: se você se sentir um pouco instável interiormente, mas tentando racionalizar as razões pelas quais a crença ainda é sólida, então provavelmente você deve encenar uma Crise de Fé. Se a crença é tão solidamente sustentada quanto a gravidade, você não precisa se preocupar — mas pense em todos os teístas que gostariam desesperadamente de concluir que Deus é tão sólido quanto a gravidade. Portanto, tente imaginar o que os céticos diriam ao seu argumento “sólido como a gravidade”. Certamente, uma razão pela qual você pode falhar em uma crise de fé é que você nunca realmente se senta e questiona em primeiro lugar — que você nunca diz: “Aqui está algo que preciso me esforçar para duvidar adequadamente”.

Se seus pensamentos estão se tornando muito complicados, é hora de você encenar uma Crise de Fé. Não faça isso de forma aleatória, não tente em um momento de distração improvisada. Não se apresse para fazê-lo rapidamente, apenas para poder dizer: “Eu duvidei porque fui obrigado a fazê-lo”. Isso não funcionaria para um teísta e também não funcionará para você. Descanse no dia anterior, para que você esteja em boas condições mentais. Reserve algumas horas ininterruptas. Encontre um lugar tranquilo para se sentar. Limpe sua mente de todos os argumentos padrão, tente ver tudo do zero. E faça um esforço desesperado para encontrar uma dúvida verdadeira que destruiria uma crença falsa e profundamente arraigada.

Elementos da técnica da Crise de Fé foram espalhados em muitos ensaios:

- [Evitar os pontos fracos reais de sua crença](#) — Uma das primeiras tentações em uma crise de fé é duvidar dos pontos mais fortes de sua crença, para que você possa [praticar](#) suas boas respostas. Você precisa procurar os pontos mais dolorosos, não os argumentos que são mais reconfortantes de considerar.

- [A Meditação sobre a curiosidade](#) — Roger Zelazny uma vez distinguiu entre “querer ser um autor” versus “querer escrever”, e também há uma distinção entre querer ter investigado e querer investigar. Não é suficiente dizer “É meu dever criticar minhas próprias crenças”; você deve ser curioso, e apenas a incerteza pode criar curiosidade. Ter em mente a Conservação da Evidência Esperada pode ajudá-lo a se [Atualizar Gradualmente](#): para cada ponto que você considera e cada elemento de novo argumento e nova evidência, você não deve esperar que suas crenças mudem mais (em média) em uma direção do que em outra - assim, você pode realmente ficar curioso a cada vez sobre como será.
- [Visão original](#) — Use a técnica de Pirsig para evitar que pensamentos padrão [armazenados em cache](#) se apressem e completem o padrão.
- A [Litania de Gendlin](#) e a [Litania de Tarski](#) — As pessoas podem suportar o que é verdadeiro, pois já estão suportando. Se uma crença for verdadeira, será melhor você acreditar nela, e se for falsa, será melhor rejeitá-la. Você aconselharia uma pessoa religiosa a tentar visualizar profundamente o mundo em que não há Deus e, sem desculpas, chegar ao pleno entendimento de que, se não houver Deus, será melhor acreditar que Deus não existe. Se alguém não consegue aceitar isso em um nível emocional profundo, não conseguirá ter uma crise de fé. Portanto, você deve fazer um esforço sincero para visualizar a alternativa à sua crença, da maneira que o melhor e mais cético gostaria que você a visualizasse. Pense no esforço que um religioso teria que fazer para imaginar, sem corromper para seu próprio conforto, a visão de um ateu sobre o universo.
- Faça um esforço extraordinário — Considere o conceito de *isshokenmei*<sup>62</sup>, o esforço desesperado e incansável para ser racional, o esforço que seria necessário para alcançar o nível de Robert Aumann e de todos os grandes cientistas ao longo da história que nunca renunciaram a suas religiões.
- [A Heurística Genética](#) (uma regra prática baseada na origem de uma ideia) — Você deve ficar extremamente desconfiado se tiver muitas ideias sugeridas por uma fonte que agora você sabe que não é confiável, mas, caramba, parece que todas as ideias acabaram dando certo. (Por exemplo, aquele que admite que a Bíblia foi escrita por mãos humanas, mas ainda se apegava à ideia de que ela contém uma sabedoria ética indispensável.)
- [A importância de dizer “opa!”](#) - Realmente é menos doloroso engolir toda a pílula amarga de uma só vez.
- [Pensamento único](#) (em contraste com o “duplipensar”, a capacidade de manter simultaneamente duas crenças contraditórias) — É o oposto do pensamento duplo. Observe os pensamentos que você evita, que surgem apenas por um momento no canto da sua mente antes de você se recusar a pensá-los. Ao tomar consciência do que você não está pensando, você pode realmente pensar.
- [Espirais de autodestruição afetiva](#) (ciclos de reforço emocional negativo) e [resistência à espiral da autodestruição feliz](#) (ciclos de reforço emocional positivo excessivo) — As espirais de autodestruição afetiva são as principais geradoras de falsas crenças de que uma Crise de Fé será necessária para se libertar. No entanto, como essas espirais também podem começar em torno de coisas reais que são genuinamente boas, você não precisa admitir que sua crença é uma mentira. Tente resistir ao efeito halo (tendência de uma impressão positiva em uma área influenciar positivamente a opinião em outras áreas) em todos os aspectos — recuse falsos elogios mesmo em relação a coisas genuinamente boas. Os [debates sobre políticas não devem parecer unilaterais](#).
- [Não proponha soluções](#) — Não apresente nenhuma solução até que o problema tenha sido discutido minuciosamente. Deixe sua mente [esperar para saber qual será sua resposta](#); e tente por pelo menos cinco minutos antes de desistir, tanto em geral quanto especialmente ao adotar o ponto de vista oposto.

E essas técnicas padrão são particularmente relevantes:

- A sequência em “[O resultado final](#)” e “[Racionalização](#)”, que explica por que é sempre errado argumentar seletivamente um lado de um debate.
- Viés positivo (tendência a favorecer informações positivas), [ceticismo motivado](#) (tendência a ser mais cético com informações indesejadas) e [parada motivada](#) (tendência a parar de buscar informações

---

62 NT.: *Isshokenmei* (一生懸命), traduzido para o português brasileiro livremente como “com todo o esforço”, é um conceito japonês que expressa a ideia de se dedicar de corpo e alma a uma tarefa, dando o seu máximo e se esforçando ao limite. É uma atitude de comprometimento total, persistência e foco, buscando a excelência em tudo o que se faz.



quando as encontradas são satisfatórias), para evitar que você procure seletivamente apoio, procure seletivamente contra-argumentos e pare seletivamente o argumento antes que ele se torne perigoso. As [alternativas ausentes](#) (falha em considerar todas as opções possíveis) são um caso especial de parada. Um caso especial de ceticismo motivado é a falsa humildade, em que você timidamente confessa que [ninguém pode saber](#) algo que você preferiria não saber. Não exija seletivamente [muita autoridade](#) dos contra-argumentos.

- Cuidado com os sinais de parada semânticos, as luzes de aplauso e sua escolha de explicar/adotar/ignorar.
- Sinta o peso dos detalhes onerosos; cada detalhe é um fardo separado, um ponto de crise.

Mas há realmente bastante material relevante aqui e no *Overcoming Bias* (Superando o Viés). A Crise de Fé é apenas o ponto crítico e o choque repentino do *isshoukenmei* mais longo — o esforço intransigente ao longo da vida para ser tão incrivelmente racional que você se eleva acima do nível de erros estúpidos. É quando você tem a chance de usar as habilidades que pratica há tanto tempo contra você mesmo.

Desejo-lhe boa sorte contra seu adversário. Tenha uma crise maravilhosa!

## 130 - O Ritual



A sala em que Jeffreyssai recebia seus visitantes não-*beisutsukai*<sup>63</sup> era silenciosamente formal, mobiliada impecavelmente com um estilo conservador. A luz do sol e o ar externo fluíam através de uma grade de prata polida, algumas arestas afiadas indicando que a parede não deveria ser aberta. O chão e as paredes eram feitos de vidro, suficientemente grosso para distorcer a visão e suficientemente profundo para ocultar o que pudesse estar abaixo. Sobre as superfícies do vidro, havia padrões delicadamente riscados sem nenhum significado específico, parecendo ter sido criados por uma criança com inclinações artísticas (o que, de fato, era o caso).

Em outras partes da casa de Jeffreyssai, havia cômodos com estilos diferentes, mas ele descobriu que a maioria dos visitantes esperava que um mestre bayesiano os recebesse dessa maneira. Ele optou por não os esclarecer; afinal, essa pequena diversão silenciosa era uma das pequenas alegrias da vida.

A convidada sentou-se em frente a ele, apoiando os joelhos em um travesseiro e os calcanhares atrás. Ela estava ali apenas em serviço de sua Conspiração, e seu traje deixava isso evidente: um macacão justo de couro rosa que cobria até as mãos — indo até o capuz que escondia sua cabeça e cabelo, embora seu rosto permanecesse simples e descoberto.

E assim, Jeffreyssai escolheu recebê-la naquela sala.

Jeffreyssai soltou um longo suspiro, exalando. “Você tem certeza?”

“Oh”, ela disse, “e eu preciso ter absoluta certeza antes que meu conselho possa mudar suas opiniões? Não é suficiente eu ser uma especialista no assunto e você não?”

Um meio sorriso apareceu nos lábios de Jeffreyssai. “Como você sabe tanto sobre as regras, afinal? Você nunca teve um treinamento formal tão extenso quanto Planck.”

“Você realmente precisa perguntar?” ela disse com um tom seco. “Se há algo que vocês *beisutsukai* adoram discutir, são as razões pelas quais vocês fazem as coisas.”

Jeffreyssai estremeceu interiormente com a ideia de buscar a racionalidade ao observar outras pessoas falando sobre isso...

“E não estremeça interiormente comigo assim”, disse ela. “Eu não estou tentando ser uma racionalista, apenas tentando ganhar uma discussão com um racionalista. Há uma diferença, como tenho certeza de que você diz aos seus alunos.”

Ela realmente consegue me ler tão bem? Jeffreyssai olhou através da grade prateada, para a luz do sol refletida na encosta facetada da montanha. Sempre, sempre a luz dourada do sol caía todos os dias, neste lugar bem acima das nuvens. Aquela luz era algo imutável. O distante Sol que aquela luz representava havia

---

63 NT: *Beisutsukai* (別つ会), em japonês, significa literalmente “encontro de despedida” ou “festa de despedida”. *Beisutsukai* é um costume japonês que consiste em uma festa ou encontro de despedida, organizado para celebrar e desejar boa sorte a alguém que está partindo, seja por mudança de emprego, transferência, fim de um curso ou qualquer outro motivo que implique em uma separação. É uma oportunidade para amigos, colegas e familiares expressarem seus sentimentos, lembrarem bons momentos e fortalecer os laços, mesmo com a distância. Nesse contexto, “visitantes não-*beisutsukai*” seriam os visitantes que não estariam fazendo uma visita com o propósito de despedida.

queimado por cinco bilhões de anos, mas agora, neste momento, o Sol ainda brilhava. E isso nunca poderia mudar. Por que desejar que as coisas permaneçam sempre iguais, quando esse desejo já foi concedido tão absolutamente quanto qualquer desejo poderia ser? O paradoxo da permanência e da impermanência: somente nessa última perspectiva havia progresso ou perda.

“Você sempre me deu bons conselhos”, disse Jeffreyssai. “Imutável, tem sido assim durante todo o tempo que nos conhecemos.”

Ela inclinou a cabeça em reconhecimento. Era verdade, e não havia necessidade de explicar as implicações.

“Então”, disse Jeffreyssai. “Não é para discutir. Só porque quero saber a resposta. Você tem certeza?” Ele nem mesmo conseguia imaginar como ela poderia adivinhar.

“Tenho certeza”, ela respondeu. “Estamos coletando estatísticas há muito tempo, e em novecentos e oitenta e cinco em mil casos como o seu...”

Então ela riu do olhar em seu rosto. “Não, estou brincando. Claro que não tenho certeza. Essa é uma coisa que só você pode decidir. Mas tenho certeza de que você deveria sair e fazer o que quer que seu povo faça - tenho certeza de que você tem um ritual para isso, mesmo que não o discuta com estranhos - quando pensa seriamente em abandonar uma premissa mantida por tanto tempo em seu relacionamento.”

Era difícil argumentar contra isso, refletiu Jeffreyssai, especialmente quando um especialista do domínio afirmava que provavelmente ele estava errado.

“Eu admito”, disse Jeffreyssai. A frase saiu de seus lábios com uma determinação final. Não havia mais necessidade de discutir: ela tinha vencido.

“Ah, pare com isso”, disse ela. Ela se levantou do travesseiro com um único movimento, sem desperdiçar um único gesto. Ela não escondia sua idade, mas também não a ostentava. Ela pegou a mão dele e a levou aos lábios para um beijo formal. “Adeus, sensei.”

“Até a próxima?” repetiu Jeffreyssai. Era uma despedida mais significativa do que um simples adeus. “Planejo visitá-la novamente, milady, e você sempre será bem-vinda aqui.”

Ela caminhou em direção à porta sem responder. Na porta, parou sem se virar. “Não será a mesma coisa”, disse ela. E então, com movimentos que não pareciam apressados, ela se afastou tão rapidamente que quase desapareceu.

Jeffreyssai suspirou. Mas pelo menos, até o desafio propriamente dito, todas as suas ações estavam prescritas, quantidades conhecidas.

Saindo da área de recepção formal, Jeffreyssai adentrou sua arena e enviou mensageiros aos seus alunos, informando que as aulas do dia seguinte seriam improvisadas durante sua ausência e que haveria um teste posteriormente.

Depois disso, ele não fez nada em particular. Passou o tempo lendo mais cem páginas do livro que havia emprestado, que não era ótimo, assim como o livro que ele havia emprestado em troca. Ele vagou pelos cômodos de sua casa ociosamente, verificando se algo havia sido roubado (apenas um baralho de cartas estava faltando). De tempos em tempos, seus pensamentos se voltavam para o desafio do dia seguinte, mas ele os deixava vagar sem direção, bloqueando qualquer conclusão ou pensamento sobre a direção de seus pensamentos.

O sol se pôs e ele o observou, com a mente em um estado de tranquilidade cuidadosamente cultivada. Era um equilíbrio delicado deixar sua mente ociosa sem ficar obcecado com isso ou gastar energia para mantê-la assim, mas anos de prática o levaram à perfeição.

Na manhã seguinte, ele acordou com a fresca lembrança do caos de seus sonhos noturnos. Fazendo o possível para preservar a sensação do caos e a memória dos sonhos, ele desceu uma escada após a outra até chegar ao quarto menos elegante de sua casa.

Ele era branco. Essa era basicamente a essência do esquema de cores.

Ao longo de uma parede, havia placas que, seguindo o método clássico e recomendado, um jovem Jeffreyssai havia escrito com extremo cuidado, gravando os conceitos em sua mente a cada toque do pincel que desenhava as palavras. “Aquilo que pode ser destruído pela verdade deve ser destruído. As pessoas conseguem suportar a verdade, pois já a suportam. A curiosidade busca sua própria aniquilação.” Até mesmo uma pequena placa que exibia apenas uma barra horizontal vermelha. Os símbolos poderiam representar qualquer coisa, revelando uma flexibilidade visual que até a Conspiração dos Bardos hesitaria em admitir abertamente.

Abaixo das placas, havia duas séries de marcas riscadas na parede. Na coluna do “mais”, duas marcas. Na coluna do “menos”, cinco marcas. Ele havia entrado nessa sala sete vezes: cinco vezes decidira não mudar de ideia e duas vezes saíra como uma pessoa diferente. Não havia proporção definida nem intervalo predefinido — isso seria uma mera zombaria. Mas se, após algum tempo, não houvesse marcas na coluna do “mais”, você poderia admitir que não fazia sentido continuar no quarto, pois você não possuía a habilidade que ele representava. Ou isso, ou você nasceu sabendo a verdade e o direito a tudo.

Jeffreyssai sentou-se não de frente para as placas, mas de costas para elas, na parede branca e inexpressiva. Era melhor não ter distrações visuais. Em sua mente, ele primeiro ensaiou o meta mnemônico e, em seguida, os vários sub-mnemônicos referenciados, para os sete princípios fundamentais e as sessenta e duas técnicas específicas que eram mais prováveis de serem necessárias no Ritual de Mudança de Mente. Além disso, Jeffreyssai acrescentou outro mnemônico, recordando seus quatorze descuidos mais embaraçosos.

Ele não inspirou profundamente. Respirar regularmente era melhor. E então, ele fez a “pergunta” a si mesmo.

