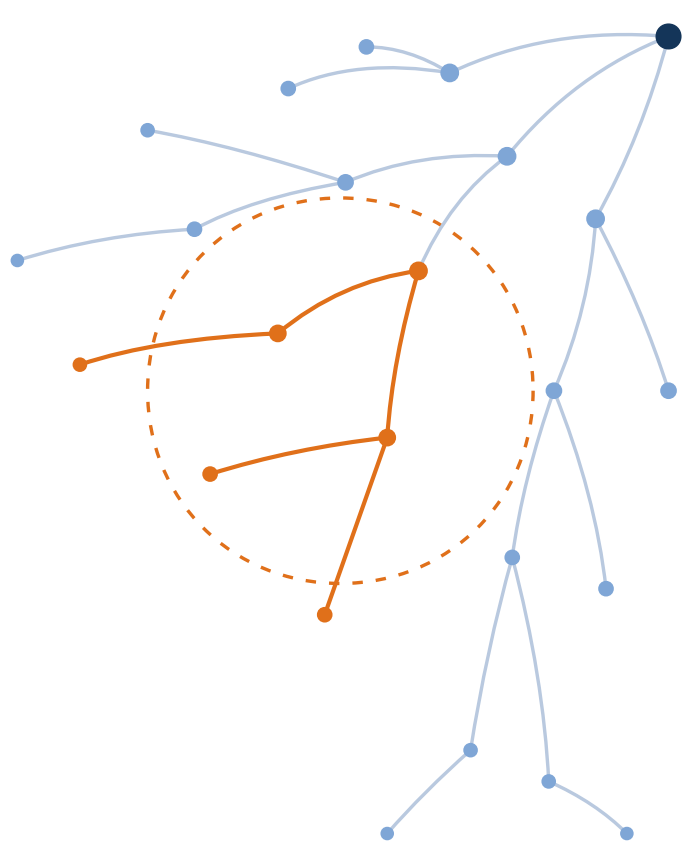




ANÁLISE DO COMPORTAMENTO

Uma análise operante de

discursos de ódio e desinformação

Márcio Moreira

Cristiano Coelho

Frederico Neves Condé

Carlos Augusto de Medeiros

Diogo Conque Seco-Ferreira

Lucas Ferraz Córdova

FICHA CATALOGRÁFICA

26-408023.0 · CDD-150

Dados Internacionais de Catalogação na Publicação (CIP) (Câmara Brasileira do Livro, SP, Brasil)

Uma análise operante de discursos de ódio e desinformação [livro eletrônico] /
Márcio Borges Moreira ... [et al.]. -- Brasília, DF : Instituto Walden4, 2026.

PDF

Outros autores: Carlos Augusto de Medeiros, Cristiano Coelho, Diogo Conque
Seco-Ferreira, Frederico Neves Condé, Lucas Ferraz Córdova

Bibliografia.

ISBN 978-85-65721-74-5

1. Comportamento agressivo 2. Comportamento - Análise 3. Comportamento -
Aspectos psicológicos 4. Comunicação - Aspectos políticos 5. Comunicação -
Aspectos sociais 6. Desinformação 7. Fake news 8. Ódio - Aspectos sociais 9.
Política - Aspectos sociais I. Moreira, Márcio Borges. II. Medeiros, Carlos Augusto
de. III. Coelho, Cristiano. IV. Seco-Ferreira, Diogo Conque. V. Condé, Frederico
Neves. VI. Córdova, Lucas Ferraz.

Índices para catálogo sistemático:

1. Comportamento : Análise : Psicologia 150

Camila Aparecida Rodrigues - Bibliotecária CRB - SP-010133/O



Instituto Walden4
Brasília, DF · 2026

A ideia deste livro

A ideia de fazermos este livro surgiu de uma postagem que Rafael Nunes Félix enviou a Carlos Augusto de Medeiros. Registramos aqui o crédito de Rafael pela ideia, com o nosso agradecimento.

Rafael Nunes Félix é psicólogo formado pelo CEUB, com interesse em pesquisa e atuação clínica na Análise do Comportamento. Foi docente na Universidade de Brasília (UnB), onde ministrou aulas de Psicologia da Personalidade sob a perspectiva da Análise do Comportamento, e hoje é aluno especial do Programa de Pós-Graduação em Ciências do Comportamento da mesma universidade. Pesquisa desconto pelo atraso e ressurgência comportamental, com foco nos processos básicos do comportamento e em suas aplicações.

Sobre os Autores

Márcio Moreira



Márcio Borges Moreira é Doutor em Ciências do Comportamento pela Universidade de Brasília, Mestre em Psicologia e Psicólogo pela Pontifícia Universidade Católica de Goiás. Professor da graduação e do Mestrado em Psicologia do Centro Universitário de Brasília (UniCEUB), onde supervisiona estágios clínicos no Centro de Formação de Psicólogos (CENFOR). É também diretor do Instituto Walden4, clínica especializada em serviços baseados em ABA para crianças e adolescentes com Transtorno do Espectro Autista (TEA).

Co-autor do livro *Princípios Básicos de Análise do Comportamento*, possui capítulos e artigos científicos publicados em periódicos nacionais e internacionais, com foco em Análise do Comportamento Aplicada, comportamento verbal e intervenções baseadas em evidências. Seus projetos de pesquisa atuais investigam a automação de treinamentos para profissionais e paraprofissionais que atuam com pessoas diagnosticadas com autismo, o uso de tecnologia e inteligência artificial no ensino de conceitos psicológicos, a formação e mudança de atitudes preconceituosas e a mensuração de comportamentos na prestação de serviços de psicologia.

Carlos Augusto de Medeiros



Dr. Carlos Augusto de Medeiros concluiu seu doutorado com louvor no programa de Ciências do Comportamento da Universidade de Brasília - UnB em 2003. Formou-se mestre em Psicologia também na UnB em 1999, tendo concluído graduação em Psicologia em 1997 na mesma instituição. Atuou como professor assistente e adjunto na Universidade Federal de Minas Gerais - UFMG de 2000-2003. No Centro Universitário de Brasília - Uni-

CEUB, de 2003 à 2024, exerceu as funções de professor de graduação, supervisor de estágio profissionalizante, coordenador, professor e orientador do curso de Mestrado em Psicologia. Ministrou diversos módulos em curso de pós-graduação em Análise do Comportamento no país, assim como atuou como supervisor de estágio clínico no Instituto Brasiliense de Análise do Comportamento - IBAC. Ao longo de sua extensa carreira, ministrou disciplinas relacio-

nadas a Análise do Comportamento, Introdução à Psicologia, Clínica Analítica Comportamental, orientou projetos de pesquisa e supervisionou estágios de pesquisa e profissionalizantes.

Desde 1997, atua como psicólogo clínico, com ênfase em terapia individual de adultos e terapia de casais. Atualmente é um dos diretores da Pragmática: Psicoterapia e Cursos, uma clínica de psicologia que oferece cursos regulares de extensão em Psicologia Clínica.

Desenvolveu um modelo de terapia analítica comportamental denominado Psicoterapia Comportamental Pragmática. Modelo este embasado na filosofia behaviorista radical e comportamento com o desenvolvimento da independência e autonomia dos clientes.

Ao longo de sua longa carreira, o Dr. Carlos Augusto de Medeiros têm se dedicado à formação de analistas do comportamento no país. É co-autor do best seller, “Princípios Básicos de Análise do Comportamento”. Possui 22 artigos, dois livros, e 43 capítulos de livros publicados. Suas publicações são relacionadas ao comportamento verbal, relações de equivalência, comportamento governado por regras, clínica comportamental e Análise do Comportamento propriamente dita.

Cristiano Coelho



Possui graduação, Mestrado e Doutorado em Psicologia pela Universidade de Brasília. Foi Professor Substituto na UnB (1999 a 2000), Professor Colaborador do Instituto de Educação Superior de Brasília - IESB (2009 a 2010), Supervisor Estadual de Pesquisas do IBGE - Goiás (2010 a 2014), Professor Colaborador do MBA em Economia Comportamental da ESPM e membro da gestão 2019-2020 da Associação Brasileira de Psicologia e Medicina Comportamental (ABPMC). Atualmente é professor Adjunto I da Pontifícia Universidade Católica de Goiás, Coordenador do Programa de Pós-Graduação Stricto Sensu em Psicologia da PUC-Goiás (Mestrado e Doutorado) e terapeuta comportamental. Tem realizado pesquisas e orientado alunos de iniciação científica, Mestrado e Doutorado principalmente nos seguintes temas: comportamento de escolha, autocontrole e impulsividade, habilidades sociais e análise do comportamento do consumidor.

Diogo Conque Seco-Ferreira



Diogo Conque Seco-Ferreira é psicólogo pela Universidade de Brasília (UnB), onde também se tornou mestre em Psicologia e doutor em Ciências do Comportamento, com pesquisa sobre comportamento do consumidor, e realizou pós-doutorado em economia comportamental na Universidade de Tilburg, nos Países Baixos. É professor associado do Departamento de Psicologia da Universidade Federal de Sergipe (UFS).

Trabalha na interface entre a Análise do Comportamento, a psicologia econômica e a economia comportamental, com pesquisas sobre o comportamento do consumidor, a análise comportamental de fenômenos sociais e as ciências comportamentais aplicadas a políticas públicas. É coautor de "Economia comportamental: uma introdução para analistas do comportamento", publicado na Revista Interamericana de Psicología.

Frederico Neves Condé



Doutor, mestre e graduado em Psicologia Social, do Trabalho e das Organizações pela Universidade de Brasília (UnB). É pesquisador-tecnologista em Informações e Avaliações Educacionais do Inep, onde chefiou o Serviço de Cálculo de Indicadores de Instrumentos Contextuais (2022–2023), foi Diretor de Avaliação da Educação Básica (2014–2015) e atuou como um dos desenvolvedores do Inse do Saeb (2019 e 2021). Foi assessor técnico da Presi-

dência Executiva e da Ouvidoria da Anatel, analista de informações sociais no MDS, consultor do PNUD e da Unesco e conselheiro no CRP-01. Atuou como docente de graduação e pós-graduação (IESB e UniCEUB) e diretor-presidente da Psicométrica Consultoria e Pesquisa. Tem como áreas de atuação a psicometria e a avaliação psicológica, educacional e social, dedicando-se à Teoria de Resposta ao Item, elaboração de instrumentos e modelagem de indicadores.

CV: <http://lattes.cnpq.br/8338146219871041>

Lucas Ferraz Córdova



Lucas Ferraz Córdova é psicólogo pela Universidade Federal de Minas Gerais (UFMG), mestre em Psicologia e doutor em Ciências do Comportamento pela Universidade de Brasília (UnB), com tese sobre o efeito de treino sucessivo no comportamento de transposição entre os operantes verbais mando e tato. É professor da Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM).

Atua em Análise do Comportamento, com ênfase em comportamento verbal, filosofia do Behaviorismo Radical e relações entre linguagem, contexto e explicação científica. Coordena projetos sobre os determinantes ambientais e históricos do comportamento de explicar, que investigam as variáveis que afetam o estabelecimento e a manutenção do repertório verbal, além de estudos experimentais sobre treino de operantes verbais, história de reforçamento e efeitos de audiências sobre o discurso explicativo.

Sumário

Apresentação	12
Alcance e limites	12
As regras de leitura	12
Como ler os quadros e os termos	13
O mapa do livro	14
Ordem de leitura	15
Referências	15
Nota sobre o processo de produção	16
O que veio dos autores	16
O que a IA fez	17
Exemplos de prompts	19
O que fizemos depois	23
Limites do processo	25
Uma leitura comportamental do próprio processo	25
Responsabilidade	25
PARTE I O fenômeno e os dados	26
Capítulo 1 — Desinformação e discurso de ódio: definições, medidas e dados	27
1.1 Definições, e o que cada uma exige medir	27
1.2 Prevalência: quanto circula e quem compartilha	29
1.3 Elites políticas como fonte	31
1.4 Posicionamento político e propensão à informação falsa: o que os estudos correlacionais mostram	32
1.5 Limitações: o que os desenhos não permitem	35
1.6 O que se sabe sobre o Brasil	36
1.7 Discurso de ódio: definições, medidas e dados	37
1.8 Interpretações rivais e limites	39
Síntese	40
Referências	41
Capítulo 2 — O que dizem as agências de checagem (e o que elas não podem dizer)	44
2.1 Duas séries de contingências	45
2.2 O que cada fonte conta	47
2.3 Por que a proporção entre as checadas não é a proporção entre as ditas	48
2.4 O que a Lupa permite dizer	49
2.5 O que o Aos Fatos permite dizer	52
2.6 Falas e menções: de quem é a falsidade	53
2.7 As agências concordam entre si?	54
2.8 O que a checagem não alcança	54
2.9 Protocolo para uma amostra comparável	56
Síntese	57
Referências	58

PARTE II Análise operante dos discursos **59**

Capítulo 3 — Declarações falsas como operantes verbais 60

3.1 A fala do político como comportamento verbal	61
3.2 A ficha de análise.....	62
3.3 Tato ou intraverbal? O contínuo e a forma de tato	65
3.4 Tato distorcido: quando a consequência do ouvinte desloca o controle pelo evento	68
3.5 Repetição: a força da afirmação repetida	71
3.6 Autoclíticos de fonte e de certeza: o que se acrescenta e o que se retira	73
3.7 A declaração como regra para a audiência	76
3.8 Casos: as mesmas quatro frases, pelo mesmo método	77
3.9 Interpretações rivais e como decidir entre elas	81
<i>Síntese</i>	82
<i>Referências</i>	83

Capítulo 4 — Atitudes políticas e preconceito como classes de equivalência . 86

4.1 Atitude como comportamento verbal avaliativo.....	87
4.2 Como uma palavra adquire função sem contato com o referente	88
4.3 Transferência de função	92
4.4 Preconceito nos experimentos	97
4.5 Atitudes em relação a presidentes do Brasil.....	101
4.6 “Nós” e “eles”: relações de oposição e de comparação.....	115
4.7 Mudança de atitude: reorganização das classes e histórias em quadrinhos.....	117
4.8 Medidas, interpretações rivais e limites	126
<i>Síntese</i>	129
<i>Referências</i>	130

Capítulo 5 — Discurso de ódio: mando disfarçado e controle aversivo 135

5.1 O que o comportamento do falante produz: aprovação, retirada de ameaças e funções que variam com o contexto.....	136
5.2 Mando disfarçado: a opinião que funciona como instrução.....	138
5.3 Controle aversivo dirigido a um grupo	141
5.4 O rótulo do grupo e a transferência de função aversiva	143
5.5 Do discurso ao ato: o que os dados permitem afirmar.....	144
5.6 Um caso com classificação externa: a fala de 2011 e o que veio depois.....	149
5.7 Interpretações rivais e limites.....	150
<i>Síntese</i>	151
<i>Referências</i>	152

**Capítulo 6 — Bolhas digitais como comunidades verbais e a esquiva de
contracontrole 155**

6.1 A bolha como comunidade verbal: audiência e fonte de reforço.....	156
6.2 O que os dados dizem sobre câmaras de eco (e onde divergem).....	158
6.3 “Realidades virtuais”: quando o controle verbal substitui o contato com o evento	162
6.4 Contracontrole e esquiva de contracontrole	163
6.5 O caso brasileiro: WhatsApp, eleições e pandemia	168
6.6 Quem curte, quem responde: uma análise exploratória de dois vídeos políticos no Instagram	171
6.7 Análises semânticas em mídias sociais	185

6.8	Interpretações rivais e limites	187
	<i>Síntese</i>	189
	<i>Referências</i>	189
Capítulo 7 — O voto como comportamento verbal		192
7.1	O voto pela definição de comportamento verbal	193
7.2	Mando, tato ou intraverbal? Funções possíveis do voto	194
7.3	O voto secreto como retirada da audiência	196
7.4	Declarar e votar: atitudes verbais e atos	198
7.5	Consequências sociais do votar: os experimentos de campo	203
7.6	Nomes e números de candidatos em classes de equivalência	205
7.7	Interpretações rivais e limites	207
	<i>Síntese</i>	208
	<i>Referências</i>	209
<hr/>		
PARTE III A ecologia da informação		210
Capítulo 8 — Economia da atenção: o algoritmo como programador de contingências		211
8.1	Atenção como reforçador	212
8.2	O que os dados mostram sobre conteúdo emocional, moralizado e negativo	214
8.3	O algoritmo como programador: quanta audiência cada emissão recebe	217
8.4	Reforçamento intermitente e engajamento	221
8.5	Experimentos com o próprio algoritmo: o que mudou e o que não mudou	226
8.6	Interpretações rivais e limites	229
	<i>Síntese</i>	230
	<i>Referências</i>	231
Capítulo 9 — O ciclo produtor-audiência		234
9.1	Opiniões podem ser reforçadas: os experimentos de 1956 a 1969	235
9.2	O que o produtor recebe: curtidas, compartilhamentos e alcance	241
9.3	O que a audiência faz e o que a mantém	243
9.4	O ciclo: cada lado modela o outro	245
9.5	Indignação moral aprendida por reforço: a evidência	246
9.6	O político como produtor de conteúdo	249
9.7	A metacontingência do ecossistema: produtor, audiência, plataforma e anunciante .	250
9.8	Interpretações rivais e limites	252
	<i>Síntese</i>	254
	<i>Referências</i>	254
Capítulo 10 — A janela de Overton: o que se pode dizer sem ser punido		257
10.1	O conceito, a origem e a tradução proposta	258
10.2	Da janela da política à janela do enunciado: o que desloca a borda	261
10.3	Informação sobre o que os outros toleram: o experimento de Pittsburgh	262
10.4	Dois eleições, duas medidas: Brasil, 2018, e Estados Unidos, 2016	265
10.5	Políticos e redes sociais: quem fala de cima e o que recebe	269
10.6	Sanção, contracontrole e o que mantém a borda no lugar	271
10.7	Como situar uma fala em relação à janela	274
10.8	Interpretações rivais e limites	275
	<i>Síntese</i>	277

Referências.....	278
------------------	-----

Capítulo 11 — Infodemiologia: como a informação se espalha e o que se mede . . . 280

11.1 Infodemiologia e infovigilância	281
11.2 Medidas de difusão: alcance, exposição, reação e a estrutura da cascata.....	283
11.3 O falso se espalha mais rápido que o verdadeiro?	286
11.4 Modelos epidemiológicos aplicados à informação	289
11.5 A infodemia de Covid-19, no mundo e no Brasil	292
11.6 O que a infodemiologia propõe fazer.....	295
11.7 Perguntas para ler um estudo de difusão.....	296
Síntese.....	297
Referências.....	298

PARTE IV Intervenções

299

Capítulo 12 — Intervenções: o que se pode mudar, em quatro níveis, e com que evidência 300

12.1 Corrigir depois: a correção, o aviso e o “efeito rebote”	302
12.2 Corrigir antes: inoculação como treino discriminativo	305
12.3 Estímulos que sinalizam precisão	309
12.4 Diante de quem fala: contra-fala, compromisso público e autoclíticos de fonte	312
12.5 Reduzir preconceito por equivalência e por histórias em quadrinhos.....	315
12.6 Mudar as contingências da plataforma.....	319
12.7 Planejamento cultural e os limites éticos de quem planeja	323
Síntese.....	326
Referências.....	327

Apêndice A — Ficha de análise de discurso 330

Apêndice B — Conceitos em uso 332

Apêndice C — Referências 336

Apêndice D — Verificação de fontes 345

D.1 Como uma afirmação chega a “verificada”	345
D.2 Números	346
D.3 Versões lidas que não são a do periódico	346
D.4 Fontes não lidas	347
D.5 Decisões de método que afetam a leitura.....	352
D.6 Deslizes de procedimento registrados	353

Apresentação

Este livro reúne uma pergunta e um método. A pergunta é o que mantém, em termos de contingências, o que políticos, plataformas e audiências dizem e repassam quando o enunciado é falso ou é classificado por alguém como discurso de ódio. O método é o da Análise do Comportamento aplicada ao comportamento verbal: em vez de inferir o que o falante tinha em mente, pergunta-se que variáveis controlam a emissão e o que ela produz em quem ouve. Skinner (1957) dá o vocabulário; a literatura experimental sobre atitudes, equivalência e contingências culturais dá o que se sabe e o que ainda não se sabe.

Foi escrito para profissionais que já conhecem a Análise do Comportamento: analistas do comportamento, psicólogos, pesquisadores. Por isso não há parte de fundamentos. Cada conceito entra, num quadro “Conceito em uso”, no ponto em que a análise passa a precisar dele, e o Apêndice B os reúne com o capítulo em que cada um foi apresentado.

Alcance e limites

Aqui não se classificam falas como falsas, não se aponta quem mente e não se defende um campo político. Quem classifica é uma agência de checagem, uma decisão judicial, uma plataforma ou um estudo, e o texto diz qual. Também não se testou nenhuma hipótese: foram lidos estudos de terceiros e, onde se propõe uma leitura operante, diz-se que é leitura. O que se tira daqui é um modo de ler o dado e de formular a pergunta que falta, e não uma resposta sobre quem é mais desonesto.

Duas consequências aparecem em todos os capítulos. A primeira é que a classificação de um operante verbal (mando, tato, intraverbal, tato distorcido) é sempre interpretação, e nunca dado: o texto diz que variáveis de controle a sustentam e qual seria a interpretação rival. A segunda é que a cobertura é desigual. Boa parte dos estudos é de língua inglesa e dos Estados Unidos, vem sobretudo do Twitter e da Covid-19, e para o Brasil há poucos; cada capítulo diz o que isso permite e o que não permite concluir.

As regras de leitura

Cinco regras valem do começo ao fim.

A primeira é o mesmo método para todos os campos políticos. Onde se analisa o que um político disse, procura-se o caso espelhado do outro campo com o mesmo critério, e quando ele não é encontrado, diz-se que não há. A regra de amostra comparável do Capítulo 2 (seção 2.9) foi fixada antes de qualquer coleta, para que a escolha dos casos não dependesse do resultado; a coleta não foi feita, e o capítulo diz por quê.

A segunda é que a falsidade é sempre da agência que a classificou. “Falso” aparece com o nome da agência, o rótulo que ela usou e a data, e nunca como propriedade que quem escreve atribui à frase.

A terceira é que “discurso de ódio” é rótulo externo. Só se chama assim o enunciado classificado por instância externa, com a instância nomeada. Sem ela, descreve-se a função do enunciado, e não se o rotula.

A quarta é a separação entre dado, interpretação e hipótese. Toda análise de dados ou de discurso traz um quadro com as três partes: o que a fonte mostra, o que se lê nisso e o que fica como hipótese, com o que seria preciso para testá-la.

A quinta trata das fontes. Só entra como fonte o que foi lido no texto integral, e cada afirmação atribuída a uma fonte foi localizada nela e lida em contexto; as planilhas desse trabalho estão descritas no Apêndice D. Onde a fonte não pôde ser obtida por via legal, escreve-se que não foi lida, e ela não é citada como verificada. As páginas citadas são as do periódico quando o texto lido as traz; quando não traz, a citação vai sem página, e o capítulo avisa a versão que foi lida (manuscrito aceito, documento de trabalho, texto em XML).

Como ler os quadros e os termos

Há três tipos de quadro. O quadro “Dado · Interpretação · Hipótese” fecha as análises. O quadro “Conceito em uso” apresenta um conceito da literatura, com a fonte e o ponto em que ele passa a ser usado; quando um capítulo o retoma, o quadro diz “retomada” e remete ao capítulo de origem. O quadro “Definição proposta” marca o termo que é proposto aqui e que nenhuma fonte lida usa, como a janela de uma audiência (Capítulo 10) e a distinção entre alcance, exposição e reação (Capítulo 11); “esquiva de contracontrole” (Capítulo 6) é proposta feita aqui, a partir de Delprato (2002), e vem sempre como hipótese.

As citações seguem o padrão APA: “Autor (ano, p. X)”. As referências de cada capítulo estão consolidadas no Apêndice C. No texto, “foi lido” e “não foi lido” querem dizer exatamente isto: o texto integral foi ou não foi lido por quem escreveu.

Os estudos citados trazem estatísticas, e o texto as explica na primeira vez em que as usa; este parágrafo é para ter à mão. O valor de p é a probabilidade de se obter uma diferença tão grande quanto a observada, ou maior, se na verdade não houvesse diferença; por convenção, p abaixo de 0,05 faz a diferença ser chamada de “significativa”, o que informa que ela dificilmente é acaso e não que seja grande ou importante. t e F são as estatísticas dos testes que produzem o p . O d e o h de Cohen medem o tamanho de um efeito, a diferença entre médias no primeiro e entre proporções no segundo, e valores perto de 0,2 são pequenos. A razão de chances (*odds ratio*, OR) compara a chance de um desfecho em dois grupos: 1 significa chances iguais. O desvio-padrão é a variação típica de uma medida, e o intervalo de confiança de 95% é a faixa de valores plausíveis para a estimativa. O R^2 é a fração da variação dos dados que um modelo explica, e a acurácia, a precisão, a revocação e o $F1$ descrevem os acertos de um classificador de texto. Quando um estudo não tem grupo de controle, ou mede só o que as pessoas declaram, o texto diz.

O mapa do livro

O livro se divide em quatro partes e quatro apêndices. A Parte I, “O fenômeno e os dados”, tem dois capítulos. O Capítulo 1 reúne as definições e os estudos de prevalência e correlacionais, com as limitações de cada medida, e fixa a regra de que o que se mede é função do instrumento. O Capítulo 2 trata das agências de checagem: o que a proporção de falsas publicada por uma agência mede e o que não mede, e por que não se compara com a de outra sem as ressalvas do capítulo.

A Parte II, “Análise operante dos discursos”, tem cinco. O Capítulo 3 é o núcleo: apresenta a ficha de análise e os operantes verbais que os seguintes usam, e a aplica a quatro falas checadas. O Capítulo 4 trata das atitudes e do preconceito como classes de equivalência. O 5 trata do discurso de ódio como mando disfarçado e controle aversivo; o 6, das bolhas digitais como comunidades verbais e da esquiva de contracontrole; o 7, do voto como comportamento verbal.

A Parte III, “A ecologia da informação”, tem quatro. O Capítulo 8 trata da economia da atenção e do algoritmo como programador de contingências; o 9, do ciclo produtor-audiência; o 10, da janela de Overton, definida pelo que uma audiência pune; o 11, da infodemiologia e das medidas de difusão. A Parte IV tem um capítulo, o 12, sobre as intervenções que decorrem da análise, em quatro níveis, e a evidência de cada uma. Os apêndices trazem a ficha de análise (A), os conceitos em uso (B), as referências consolidadas (C) e a verificação de fontes, com as lacunas (D).

Ordem de leitura

Os Capítulos 1 e 2 podem ser lidos à parte, por quem quer só os dados e a leitura crítica das checagens. O Capítulo 3 é o ponto de apoio dos demais: quem vai ler a análise de falas e de discurso de ódio, nos Capítulos 5 e 7, deve passar por ele e pelo 4. Fora isso, os capítulos são autônomos, com remissões explícitas ao que usam de outro. O Capítulo 12 se lê melhor por último, porque usa o quadro de treino discriminativo, as distinções dos Capítulos 8 e 9 e a definição de alcance, exposição e reação do Capítulo 11.

Quem usa o livro como ferramenta de análise encontra perguntas prontas em quatro pontos: a ficha do Capítulo 3 (seção 3.2), reproduzida no Apêndice A, para ler uma fala; as do Capítulo 10 (seção 10.7), para situar uma fala numa janela; as do Capítulo 11 (seção 11.7), para ler uma medida de difusão; e as do Capítulo 12 (seção 12.7), para ler uma intervenção. A ficha é a de maior uso: transforma o que os capítulos descrevem numa sequência de perguntas sobre a fala, a audiência e as consequências, e termina sempre na que importa, que é o que mantém a fala.

Referências

- Delprato, D. J. (2002). Countercontrol in behavior analysis. *The Behavior Analyst*, 25(2), 191–200. <https://doi.org/10.1007/bf03392057>
- Skinner, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts.

Nota sobre o processo de produção

Este livro foi escrito com uso intensivo de inteligência artificial generativa, e esta nota descreve como. Os autores somos nós, os seis signatários (Márcio Moreira, Carlos Augusto de Medeiros, Cristiano Coelho, Diogo Conque Seco-Ferreira, Frederico Neves Condé e Lucas Ferraz Córdova): definimos o problema, o quadro interpretativo e os critérios de análise, e respondemos por cada afirmação publicada. A IA redigiu as versões iniciais de todos os capítulos, da Apresentação e dos Apêndices A e B, e conduziu a verificação das fontes; nós as revisamos, alteramos, ajustamos e conferimos, entre 19 e 26 de outubro de 2026. A condução da IA, do pedido original ao último pedido de revisão, foi toda feita por Márcio Moreira.

Preferimos descrever o processo em detalhe a resumi-lo numa linha de agradecimento. O leitor tem o direito de saber como o texto foi produzido para avaliá-lo, e uma descrição vaga serviria mal tanto à transparência quanto ao trabalho que de fato fizemos. Pelo mesmo motivo, não chamamos este de um livro “gerado por IA”: a expressão descreve mal um processo em que o problema, o método e o juízo final foram humanos do início ao fim.

O que veio dos autores

O que orienta o livro antecede a redação com IA e veio de nós, no pedido original de 10 de outubro de 2026 e nos pedidos seguintes.

- **O problema.** A escolha do discurso de ódio e da desinformação, com foco nas declarações de políticos brasileiros, como objeto, e a pergunta central: o que uma análise skinneriana consegue dizer sobre esse fenômeno a partir dos dados que a ciberpsicologia e as áreas vizinhas (ciência política, comunicação, infodemiologia) já produziram.
- **O quadro interpretativo.** Uma posição skinneriana explícita, fundada no comportamento verbal (*Verbal behavior*, de Skinner, e *Schedules of reinforcement*, de Ferster e Skinner, que abrimos como exceção à regra de que só se citam artigos), combinada com a análise contextual de Bernard Guerin. Do artigo de Guerin que indicamos como base entram as atitudes e as crenças como comportamento verbal mantido por comunidades verbais, as funções de tato, intraverbal e mando de um mesmo enunciado e o papel dos autoclíticos na negociação social do que se diz; de outro trabalho dele, a análise do conhecimento como construção social e a condição em

que um grupo pequeno e coeso sustenta o que diz. Somam-se a isso a equivalência de estímulos e a transferência de função, trabalhadas em livros nossos sobre atitudes e sobre histórias em quadrinhos. Esse quadro chegou às ferramentas pelo pedido original, pelo artigo-base, por esses livros e pelas cópias de Skinner que fornecemos.

- **A seleção da literatura.** Fixamos as regras: só artigos científicos, mais nossos livros sobre atitudes e histórias em quadrinhos e as duas obras de Skinner; só textos obtidos por via legal, sem Sci-Hub nem equivalentes; só se cita o que foi lido no texto integral; sites que proíbem rastreadores de IA, como os da Lupa e do Aos Fatos, não são coletados em massa. A lista de candidatas veio em parte de nós (27 artigos do nosso acervo e 12 livros nossos) e em parte da IA (97 artigos previstos a partir do sumário e 34 acrescentados depois de 20 buscas dirigidas no OpenAlex). Aprovamos o levantamento em 10 de outubro de 2026, e ele chegou a 168 fontes, das quais 108 foram lidas no texto integral e 60 ficaram pendentes (Apêndice D.2 e D.4).
- **Os critérios de análise.** O que conta como contingência, como consequência social e como controle verbal. Classificar um operante verbal (mando, tato, intraverbal, tato distorcido) é sempre interpretação, e nunca dado. O mesmo método vale para políticos de todos os campos; “falso” é sempre a classificação de uma agência nomeada; “discurso de ódio” é rótulo de uma instância externa; dado, interpretação e hipótese vêm separados. E o que não pretendemos afirmar: não oferecemos dados robustos sobre o fenômeno, mas leituras que possam orientar experimentos futuros em Análise do Comportamento.
- **A estrutura.** Os dez tópicos do pedido original (dados e estudos correlacionais, agências de checagem, operantes verbais, equivalência de estímulos, bolhas digitais, voto, economia da atenção, ciclo produtor-audiência, infodemiologia e intervenções), o capítulo sobre a janela de Overton, que acrescentamos em 11 de outubro, e a aprovação do sumário em quatro partes, em 10 de outubro.

O que a IA fez

Usamos o Claude Sonnet 5.5, da Anthropic, no Claude Code, ambiente em que o modelo lê e escreve os arquivos do projeto, executa scripts e consulta bases bibliográficas abertas, entre 10 e 26 de outubro de 2026. O histórico de versões mostra também o Claude Opus 5.5 em tarefas de organização do repositório e

de diagramação, sem redação de texto do livro. As ferramentas trabalharam sempre a partir do quadro interpretativo e das instruções que fornecemos, e quem as conduziu, do começo ao fim, foi Márcio Moreira.

- **Levantamento e obtenção das fontes.** Propôs candidatas, conferiu autores, ano, periódico, volume e páginas no Crossref, obteve os textos por vias legais (acesso aberto, repositórios, versões postadas pelos autores e os PDFs que anexamos) e extraiu o texto, com reconhecimento óptico de caracteres quando a camada de texto era ruim. O que não obteve ficou numa lista de pendências e não é citado como verificado.
- **Estrutura.** Propôs o sumário detalhado e a ordem de escrita: os capítulos 4, 9, 6, 8 e 10 primeiro, porque tinham as fontes completas, e os demais em seguida.
- **Redação inicial.** A partir do quadro, desenvolveu as primeiras versões de todas as seções teóricas, da Apresentação e dos Apêndices A e B.
- **Aplicação do quadro aos estudos.** Produziu as primeiras leituras skinnerianas dos estudos selecionados: descrição dos procedimentos, identificação de possíveis contingências, interpretações rivais e limites, e a interpretação dos resultados, sempre com dado, interpretação e hipótese separados.
- **Verificação das afirmações.** Registrou cada afirmação atribuída a uma fonte numa planilha, com as frases literais que a sustentam; um script procurou essas frases no texto integral, e a IA leu o contexto e registrou o resultado e as ressalvas. Um segundo script conferiu que toda citação do texto tem referência e toda afirmação marcada existe e está verificada. A planilha tem 1.207 entradas (Apêndice D.1). No levantamento inicial, a conferência corrigiu autores, anos de volume e páginas que a IA havia registrado de memória e fez retirar uma afirmação atribuída a um artigo que ele não faz.
- **Figuras, referências e arquivos.** Gerou as figuras por script e recortou, dos livros do Instituto Walden⁴, as imagens dos estudos que o Capítulo 4 descreve; montou as referências em APA e os arquivos de entrega (.docx, HTML e PDF).
- **Reescritas a pedido.** Reescreveu em prosa as descrições de estudos e refez o Capítulo 4 com as imagens dos nossos estudos e das histórias em quadrinhos, conforme os pedidos descritos abaixo.

A IA não é autora. Decidiu, por delegação nossa, muitas questões de redação (a ordem dos capítulos, a numeração, o título de capítulos, o tratamento de lacunas, o que ficou de fora quando uma fonte não pôde ser obtida), e cada decisão está registrada com data e com o caminho para revertê-la. Três termos que

aparecem como definições próprias do livro nasceram na redação com IA: “esquiva de contracontrole” (Capítulo 6), “janela de uma audiência” (Capítulo 10) e a distinção entre alcance, exposição e reação (Capítulo 11). Cada um é declarado como proposta no texto, e a avaliação deles é nossa. A palavra final sobre cada interpretação é dos autores.

A redação se deu em sessões de conversa com a IA sobre os arquivos do projeto, um capítulo por vez, a partir de 10 de outubro de 2026. Começou pelo pedido original, que o repositório guarda sem alterações. A IA entregou por etapas (o artigo-base, o sumário, o levantamento bibliográfico e só então os capítulos), e aprovamos cada etapa antes da seguinte. Em outros momentos respondemos “você pode decidir por mim” e “Continue a redação do livro e a finalize. Você pode tomar todas as decisões”: a IA decidiu e registrou cada decisão, com data, num arquivo próprio, e a delegação não valeu como aprovação de capítulo. Pedimos também correções de rumo depois de ler o rascunho: que o texto deixasse de tratar “o livro” como quem lê, diz e propõe; que as aberturas de seção se bastassem sem o título; que as descrições de pesquisa, truncadas e pouco didáticas, fossem revistas parágrafo a parágrafo; e que o Capítulo 4 fosse refeito com figuras coloridas e didáticas e com as imagens dos nossos estudos, inclusive das histórias em quadrinhos. Alguns desses pedidos aparecem textualmente na próxima seção. O registro completo, com o pedido original, as decisões datadas, a planilha de verificação e o histórico de versões, está no repositório do projeto e pode ser consultado mediante solicitação.

Exemplos de prompts

Quase todos os pedidos foram mensagens curtas, em linguagem corrente, feitas no meio do trabalho; a exceção é o pedido original, longo e estruturado, que reproduzimos na íntegra. Todos foram escritos por Márcio Moreira. Os demais exemplos são textuais e vêm de pontos diferentes do processo; neles a grafia está corrigida, e as palavras acrescentadas ou trocadas estão entre colchetes. Cada exemplo é seguido do que a IA fez com ele, segundo o registro do projeto.

O pedido original (10 de outubro de 2026), na íntegra. Os títulos das partes do pedido aparecem em negrito.

PAPEL

Você é um analista do comportamento (referencial skinneriano, Análise Experimental e Aplicada do Comportamento) com experiência em comportamento verbal, equivalência de estímulos e análise de práticas culturais, escrevendo para um público de graduação avançada e pós-graduação em Psicologia.

OBJETIVO

Escrever o livro “Uma análise operante de discursos de ódio e desinformação”, em português do Brasil, com linguagem didática e rigorosamente técnica (terminologia da Análise do Comportamento, definições operacionais, sem mentalismo explicativo). O foco são discursos e declarações falsas de políticos brasileiros.

FLUXO DE TRABALHO (execute em etapas e pare para minha aprovação ao fim de cada uma)

1. Leia primeiro o artigo-base: <https://pmc.ncbi.nlm.nih.gov/articles/PMC2733686/> e resuma como ele fundamentará a análise de comportamento verbal.
2. Proponha o sumário detalhado (capítulos, seções, figuras previstas, fontes previstas).
3. Monte o levantamento bibliográfico (ver “Regras de fontes”) e a tabela de verificação, antes de escrever qualquer capítulo.
4. Escreva um capítulo por vez, com figuras, e entregue como documento.
5. Ao final, faça a revisão cruzada de todas as citações e referências.

CONTEÚDO (acrescente tópicos pertinentes que eu não listei)

- a) Dados e estatísticas sobre desinformação (“fake news”) e discurso de ódio, incluindo estudos correlacionais entre propensão à desinformação e posicionamento político. Apresente limitações (correlação $\neq$ causa, amostras, medidas, contexto cultural).
- b) Dados de agências de checagem (Aos Fatos, Lupa, etc.) sobre declarações falsas ou enganosas de Jair Bolsonaro durante o mandato e de Lula. Use o mesmo critério para os dois: mesma agência, período comparável, mesma metodologia de classificação, denominador explícito (nº de declarações checadas) e ressalva sobre viés de seleção das checagens.
- c) Análise dos discursos em termos dos operantes verbais (mando, tato, ecoico, intraverbal, textual, etc., e tato distorcido/mando disfarçado), com base no artigo-base.
- d) Análise em termos de equivalência de estímulos: formação de classes, transferência de função, discriminações condicionais, propriedades (reflexividade, simetria, transitividade, equivalência), relação com formação de preconceito e atitudes. Referencie os livros de Márcio Moreira sobre atitudes e sobre HQs e atitudes (walden4.com.br/livros).
- e) Bolhas digitais como comunidades verbais; criação de “realidades virtuais” por controle verbal; esquivas de contracontrole.
- f) Comportamento de votar como comportamento verbal.
- g) Economia da atenção; pesquisas sobre a tendência de usuários a consumir conteúdo emocional/moralizado; algoritmos como programas de reforçamento.

h) Ciclo de reforçamento entre produtor de conteúdo e audiência (curtidas, compartilhamentos, audiência como reforçadores generalizados), análise das contingências de ambos.

i) Epidemiologia da informação (infodemiologia): difusão, taxas de propagação, intervenções.

j) Implicações práticas: intervenções culturais e educacionais baseadas em análise do comportamento.

FIGURAS

Ilustre todo procedimento ou conceito que se beneficie de diagrama: discriminações condicionais (MTS), classes de equivalência e suas propriedades, tríplique contingência, ciclo produtor-audiência, bolhas como comunidades verbais. Gere as figuras como arquivos (Graphviz/matplotlib/SVG), com legenda numerada, vetoriais, legíveis em impressão P&B e originais (não reproduzir figuras de terceiros).

REGRAS DE FONTES

- Cite apenas artigos científicos, mais seus livros sobre atitudes e HQs (exceção explícita). Não cite outros livros.
- Obtenha os artigos apenas por vias legais: acesso aberto (PMC, SciELO, Unpaywall, OpenAlex, repositórios), acesso institucional que eu disponibilizar, ou PDFs que eu anexar. Não use Sci-Hub ou equivalentes. O que não for obtido legalmente entra em uma lista “PDF pendente” e NÃO é citado como verificado.
- Para cada artigo citado, leia o texto completo e confirme que ele contém a informação atribuída. Registre numa tabela: referência | DOI | trecho/página que sustenta a afirmação | status (verificado / pendente).
- Valide DOI, autores, ano, periódico, volume e páginas (ex.: Crossref). Nunca invente ou complete referência de memória; se não conseguir verificar, marque como pendente.
- Faça duas passagens de verificação (citações no texto ↔ lista de referências ↔ tabela de verificação). Use formato APA 7.
- Citações literais: curtas, entre aspas, com página, e acompanhadas de imagem (recorte) do trecho do PDF original. Prefira paráfrase fiel.

NOTÍCIAS

Use matérias apenas de G1, CNN Brasil, BandNews e Canal Meio. Para cada uma: trecho curto citado com crédito, URL, data de publicação e data de acesso, mais captura de tela do trecho. Registre também a data de cada fala analisada e a checagem correspondente.

TOM E RIGOR POLÍTICO

Descreva comportamentos e contingências, não “essências” de pessoas ou partidos. Analise políticos de diferentes campos com o mesmo método. Separe sempre dado verificado, interpretação analítica e hipótese. Evite afirmações difamatórias; atribua cada classificação de falsidade à agência que a fez.

ENTREGA

Cada capítulo como documento (.docx), figuras em pasta separada, tabela de verificação de referências em planilha, e lista de pendências (PDFs não obtidos, dados não confirmados).

O que a IA fez. Desse pedido vieram o trabalho em etapas (artigo-base, sumário, levantamento bibliográfico, capítulos e revisão cruzada), a planilha de verificação, a regra de só citar o que foi lido no texto integral e a separação entre dado, interpretação e hipótese em todos os capítulos. A instrução de figuras legíveis em preto e branco foi revista depois, a nosso pedido: as figuras do Capítulo 4 passaram a coloridas em 16 de outubro, e as demais em 26 de outubro de 2026.

Aprovar e seguir (10 de outubro de 2026). Entre as etapas, respondemos com frases como estas:

Vá em frente.

Aprovado. Continue.

O que a IA fez. Tratou cada resposta como aprovação só da etapa que a pergunta cobria; os pontos que não foram respondidos continuaram em aberto no registro de decisões.

Delegar decisões (10 e 14 de outubro de 2026).

Você pode decidir por mim.

Continue a redação do livro e a finalize. Você pode tomar todas as decisões.

O que a IA fez. Decidiu o que faltava (ordem de escrita, numeração, tratamento de fontes que não foram obtidas, títulos) e registrou cada decisão, com data e com o caminho para revertê-la. A delegação não valeu como aprovação de capítulo.

Acrescentar escopo (11 de outubro de 2026).

Adicione um capítulo com análise comportamental da janela de Overton, relacionando com discursos de políticos e redes sociais.

O que a IA fez. Escreveu o capítulo, hoje o Capítulo 10, e propôs a definição de “janela de uma audiência” que ele usa, declarada como proposta no texto.

Responder a uma pergunta da IA, com uma resposta ambígua (11 de outubro de 2026). A IA tinha perguntado se a exceção de fontes aberta para Skinner valia também para a obra de Ferster e Skinner sobre esquemas de reforçamento. A resposta foi:

Fester, ok. Vá em frente.

O que a IA fez. Leu “Fester” como Ferster e Skinner, registrou essa leitura e avisou que, se o livro pretendido era outro, bastava dizer.

Corrigir a redação depois de ler o rascunho (16 de outubro de 2026).

“O livro” entr[a] como uma entidade no próprio livro (e.g., “O primeiro estudo que o livro leu é um jogo”). Há parágrafos de início de seção que começam de forma abrupta e confusa ou considerando que o título da seção já [faz] parte do parágrafo. Identifique e ajuste essas questões.

O que a IA fez. Reescreveu cerca de 185 frases em voz impessoal, em todos os capítulos, e refez as aberturas de seção para que se entendam sem o título.

Há descrições de pesquisa que estão truncadas, pouco didáticas. Revise parágrafo a parágrafo essa questão da redação.

O que a IA fez. Reescreveu em prosa as descrições de estudos dos doze capítulos e da Apresentação, uma por uma, com a pergunta, a amostra, o procedimento, a medida, o resultado e a ressalva de cada estudo.

Pedir detalhe e imagens (16 de outubro de 2026).

No capítulo de atitudes e equivalência, use em detalhes e usando as imagens desses dois trabalhos: [endereços dos dois trabalhos]. Use figuras coloridas, muito didáticas. Refaça o capítulo com este tom.

Insira, onde for possível e adequado, referências a: [endereços de onze trabalhos]. Use imagens desses livros, deixe tudo bem ilustrado, incluindo imagens das HQs.

O que a IA fez. Refez o Capítulo 4 com figuras coloridas, recortou imagens dos estudos e das histórias em quadrinhos do Instituto Walden4 e incluiu as obras nas referências, com o endereço de cada PDF. Um dos endereços aparecia duas vezes na lista, e a IA o leu como o volume seguinte, registrando a suposição para que a confirmássemos.

O que fizemos depois

Entre 19 e 26 de outubro de 2026, os autores trabalharam sobre o texto produzido nas etapas anteriores: revisaram, alteraram, ajustaram e conferiram os capítulos, a Apresentação e os apêndices, num manuscrito comum.

1. **Revisão conceitual.** Avaliamos se cada argumento era coerente com o quadro skinneriano e com a análise contextual, e descartamos ou refizemos os que não eram.

2. **Alteração e ajuste.** Corrigimos erros conceituais, inferências sem apoio nos dados e extrapolações, e reescrevemos os trechos que pediam outra formulação.
3. **Conferência.** A conferência de cada afirmação contra o texto integral da fonte foi feita primeiro pela IA, com o procedimento do Apêndice D; nós conferimos os textos nessa mesma semana de revisão. As 60 fontes que não pudemos obter não são citadas como verificadas, e onde o assunto exigiria a fonte o capítulo diz que ela não foi lida.
4. **Edição final.** Unificamos terminologia, estilo e a articulação entre capítulos.

O quadro a seguir resume as etapas por capítulo.

QUADRO

Etapas do trabalho, por capítulo

Capítulo	Primeira versão pela IA	Obras na lista do capítulo	Entradas verificadas	Revisão e conferência pelos autores
Apresentação	15/10/2026	2	0	19 a 26/10/2026
1. Definições e dados	14 a 15/10/2026	21	48	19 a 26/10/2026
2. Agências de checagem	10/10/2026	10	39	19 a 26/10/2026
3. Declarações falsas	14 a 15/10/2026	13	92	19 a 26/10/2026
4. Atitudes e equivalência	10/10/2026, refeito em 16/10/2026	39	300	19 a 26/10/2026
5. Discurso de ódio	14 a 15/10/2026	13	45	19 a 26/10/2026
6. Bolhas digitais	11/10/2026	13	87	19 a 26/10/2026
7. O voto	14 a 15/10/2026	13	54	19 a 26/10/2026
8. Economia da atenção	11/10/2026	12	80	19 a 26/10/2026
9. Ciclo produtor-audiência	10/10/2026	13	123	19 a 26/10/2026
10. Janela de Overton	11/10/2026	11	80	19 a 26/10/2026
11. Infodemiologia	14 a 15/10/2026	12	97	19 a 26/10/2026
12. Intervenções	14 a 15/10/2026	27	162	19 a 26/10/2026

Nota. As datas da primeira versão são as do primeiro registro de cada arquivo no histórico do projeto; o Capítulo 10 foi começado em 11/10/2026 com outro número e renumerado em 14/10/2026. “Entradas verificadas” são as afirmações e os dados de fontes registrados na planilha

de verificação para o capítulo, cada um com o trecho lido em contexto, antes da revisão dos autores e da seção 6.7. Em 16/10/2026 pedimos a todos os capítulos duas revisões de redação (o texto sem “o livro” como agente e as descrições de estudos em prosa), e o Capítulo 4 foi refeito. Os Apêndices C e D são gerados por script a partir da planilha.

Limites do processo

Dois limites merecem ser ditos. O primeiro é que a verificação confirma que a passagem atribuída a uma fonte está no texto dessa fonte, na página indicada, e não que a leitura operante que se faz dela esteja certa; essa é a parte que os capítulos chamam de interpretação ou de hipótese, e a que mais depende do nosso juízo. O segundo é que modelos de linguagem erram de memória, e a própria IA registrou, no início do trabalho, dados que o Crossref e o texto das fontes desmentiram; por isso nada entrou no livro sem a conferência descrita no Apêndice D, e onde a fonte não pôde ser lida o texto diz que não foi lida.

Uma leitura comportamental do próprio processo

Para analistas do comportamento, a pergunta sobre autoria tem uma resposta conhecida, e ela não muda por causa da IA. Num ensaio de 1972, “On ‘having’ a poem”, Skinner compara escrever um poema a ter um filho: o poeta não cria a partir do nada, é o lócus onde variáveis genéticas e ambientais se encontram. O que chamamos de criatividade é variação e seleção, e boa parte do trabalho do escritor está em editar, isto é, em selecionar o que permanece.

Não invocamos Skinner para diminuir o papel da IA nem para inflar o nosso. A analogia vale para descrever o que de fato mudou: a IA ampliou muito a fonte de variação, e o trabalho autoral se concentrou ainda mais no arranjo das condições (o problema, o quadro, as instruções) e na seleção (o que foi aceito, corrigido, reescrito ou descartado). Essa é uma descrição do processo, não uma justificativa dele; se o leitor discordar dela, a descrição das seções anteriores continua valendo por si.

Responsabilidade

Os autores revisaram o conteúdo deste livro e assumem responsabilidade por ele, inclusive pelas partes cuja primeira redação foi feita com IA. Erros remanescentes são nossos.



PARTE I

O fenômeno e os dados

- 1 Desinformação e discurso de ódio: definições, medidas e dados
- 2 O que dizem as agências de checagem (e o que elas não podem dizer)

1

Desinformação e discurso de ódio: definições, medidas e dados

Os números de Allcott e Gentzkow (2017) sobre a eleição presidencial americana de 2016 mostram, logo de saída, o problema que organiza este capítulo. No banco de notícias falsas que reuniram, 115 favoráveis a Trump foram compartilhadas 30 milhões de vezes no Facebook, e 41 favoráveis a Clinton, 7,6 milhões. Os mesmos autores estimam que o adulto americano médio viu e lembrou 1,14 notícia falsa e dizem, de forma expressa, que não avaliam se elas decidiram a eleição. Um leitor apressado junta esses números numa frase, e eles respondem a perguntas diferentes: o primeiro mede o que circulou; o segundo, o que alguém viu e lembrou; nenhum mede o que mudou no que alguém fez.

Este capítulo sistematiza o conhecimento acumulado sobre quantidade e forma de circulação da desinformação e do discurso de ódio, bem como o perfil de quem os produz e compartilha, explicitando as métricas empregadas e seus limites metodológicos. É o capítulo menos analítico do livro: não propõe leitura operante, e sim mostra os dados com que os capítulos seguintes trabalham e o que cada um deles exige que se diga em voz baixa. Vale nele a regra de todo o livro: antes de ler o número, perguntar o que o instrumento controla.

1.1 Definições, e o que cada uma exige medir

Cada definição de “informação falsa” traz embutida uma exigência de medida. Lazer et al. (2018) definem notícia falsa como informação fabricada que imita o conteúdo jornalístico na forma, mas não no processo organizacional nem na intenção; é uma subcategoria da desinformação em sentido amplo (misinformation), isto é, informação incorreta ou enganosa sobre o estado do mundo, e

seus veículos carecem das normas e dos processos editoriais da mídia. Se o elemento definidor é a intenção e o processo do editor, escrevem eles, a prevalência deve ser medida por editor, e não por notícia individual. Allcott e Gentzkow (2017) definem notícia falsa como artigo de notícia intencional e verificavelmente falso, que pode enganar o leitor, e excluem da definição erros não intencionais de reportagem. As duas definições exigem, em princípio, conhecer a intenção; nenhuma das medidas que as usam a observa.

Os estudos de compartilhamento resolvem o problema trocando o item pela fonte. Guess et al. (2019) tomam como notícia falsa o conteúdo falso ou enganoso intencionalmente vestido como notícia, em geral para gerar receita de anúncios, e cruzam os links que cada respondente compartilhou com listas de editores de notícias falsas feitas por jornalistas e acadêmicos. Lasser et al. (2022) usam a nota que o NewsGuard, um serviço de avaliação da confiabilidade de sites de notícias, dá ao domínio (de 0 a 100, abaixo de 60 “não confiável”) como estimador da confiabilidade do item compartilhado. É uma convenção explícita, a de inferir o item pela fonte, e não é um defeito escondido; mas muda o que o número diz: a proporção de links para domínios não confiáveis não é a proporção de afirmações falsas.

Os estudos experimentais trocam de novo o objeto. Rathje et al. (2023) escrevem que *misinformation* pode se referir a notícias fabricadas, boatos falsos, teorias da conspiração ou desinformação, e o que medem é o julgamento de manchetes verdadeiras e falsas pelos participantes. E Ricard e Medeiros (2020), numa nota sobre o caso brasileiro, distinguem a informação falsa que quem a dissemina acredita ser verdadeira daquela que quem a dissemina sabe ser falsa — dois pontos de um “espectro da desordem informacional”, sem apresentar mensuração empírica própria. A distinção é de intenção, e a Tabela 1.1 mostra que quase nenhuma das medidas a observa. .

Os termos são usados como as fontes os trazem. “Informação falsa” e “desinformação” aparecem, sem pretensão de definição própria, para o que a fonte chama de *misinformation*, e a intenção só entra quando a fonte a medir. Quanto a “discurso de ódio”, a regra é a do restante do livro: é um rótulo externo, aplicado só ao enunciado classificado por instância externa, com a instância nomeada; os estudos da seção 1.7 usam a expressão como categoria da própria medida, e aqui ela é tomada como eles a trazem.

TABELA 1.1*Seis definições de informação falsa e o que cada uma exige medir*

Fonte	Termo e definição	O que a definição exige medir	O que o estudo mediu
Lazer et al. (2018)	Notícia falsa: informação fabricada que imita a forma do jornalismo, não o processo nem a intenção	Processo e intenção do editor	Sem medida própria: ensaio que propõe medir por editor, e não por notícia"
Allcott e Gentzkow (2017)	Notícia falsa: artigo intencional e verificavelmente falso, que pode enganar	Intenção e veracidade, notícia a notícia	Compartilhamentos de 156 notícias classificadas como falsas por checadores nos três meses antes da eleição
Guess et al. (2019)	Conteúdo falso ou enganoso vestido como notícia, em geral por receita de anúncios	Origem do conteúdo	Links compartilhados no Facebook, cruzados com listas de editores de notícias falsas
Lasser et al. (2022)	Informação não confiável, definida pelo domínio	Confiabilidade da fonte, não do item	Nota do NewsGuard abaixo de 60 para o domínio do link
Rathje et al. (2023)	Misinformation: notícias fabricadas, boatos, teorias da conspiração ou desinformação	Veracidade da manchete	Julgamento de manchetes verdadeiras e falsas por participantes, com e sem incentivo
Ricard e Medeiros (2020)	Misinformation (o disseminador crê) e disinformation (o disseminador sabe), citando outra obra	O que o disseminador sabe	Sem medida própria: nota comentada, sem dado empírico

Nota. Descrições próprias, a partir dos trechos citados; a coluna "O que o estudo mediu" traz o que cada estudo operacionalizou, e não o que declarou querer medir.

1.2 Prevalência: quanto circula e quem compartilha

A prevalência pode ser investigada a partir de três perguntas centrais, e cada uma exige um instrumento próprio de medida: quanto circula, quem compartilha e de onde o conteúdo provém. A primeira pergunta diz respeito a quanto circula. Allcott e Gentzkow (2017) montaram, para a eleição presidencial americana de 2016, um banco com 156 notícias relacionadas à eleição que sites de checagem de referência classificaram como falsas nos três meses anteriores à votação. As 115 favoráveis a Trump foram compartilhadas 30 milhões de vezes no Facebook, e as 41 favoráveis a Clinton, 7,6 milhões de vezes. Compartilhamento, porém, não é o mesmo que ter visto, e os autores tentaram estimar o

que as pessoas viram: numa pesquisa online com 1.200 pessoas, feita depois da eleição, perguntaram se elas se lembravam de ter visto cada uma de 14 manchetes falsas que haviam circulado. Cerca de 15% disseram que sim, mas cerca de 14% também disseram ter visto manchetes-placebo, falsas, inventadas pelos autores e nunca circuladas. Como as manchetes-placebo não podiam ter sido vistas, os autores tomam a diferença entre as duas porcentagens como medida de recordação verdadeira e, projetando-a para o conjunto de notícias do banco, estimam que o adulto médio viu e lembrou 1,14 notícia falsa. Na mesma pesquisa, só 14% dos adultos americanos viam as mídias sociais como a fonte mais importante de notícias eleitorais.

A segunda pergunta é quem compartilha, e o estudo de Guess et al. (2019) é o que mais se aproxima de uma resposta com registro de comportamento, e não só com autorrelato. Os autores aplicaram uma pesquisa online a uma amostra representativa de 3.500 americanos e pediram a cada respondente que autorizasse o acesso, por um aplicativo, a dados do próprio Facebook, entre eles as postagens com links externos; a análise cobre quem consentiu e teve os dados ligados à pesquisa. Os links compartilhados em 2016 foram então comparados com listas de editores de notícias falsas feitas por jornalistas e acadêmicos, o que evita os vieses conhecidos do autorrelato de atividade online. O resultado principal é de raridade: cerca de 8,5% dos respondentes com dados ligados compartilharam ao menos uma matéria de domínio de notícias falsas, e mais de 90% não compartilharam nenhuma. Não é que as pessoas compartilhem pouco em geral: no período, 26,1% dos respondentes compartilharam de 10 a 100 links de qualquer tipo, e 61,3%, de 100 a 1.000. Entre os que compartilharam notícias falsas, havia mais republicanos: 38 contra 17 democratas em números absolutos, e 18,1% dos republicanos contra 3,5% dos democratas na amostra. O achado mais robusto, dizem os autores, é a idade: quem tinha 65 anos ou mais compartilhou em média 0,75 matérias de domínios de notícias falsas, mais que o dobro do segundo grupo etário mais velho (0,26), e a diferença persiste quando se mantêm constantes a ideologia e a educação.

Duas ressalvas dos próprios autores acompanham esses números. A primeira é que o padrão por ideologia é coerente com dois fatos: a maioria das notícias falsas tinha inclinação pró-Trump, e as pessoas tendem a compartilhar o que aprovam. O dado, portanto, pode não indicar que conservadores tenham maior tendência a compartilhar notícias falsas quando expostos a elas. A segunda é que os autores não têm medida de se os respondentes sabiam que o conteúdo era falso (Guess et al., 2019). Lazer et al. (2018) acrescentam um limite de método. Eles citam um estudo que avaliou a disseminação de um conjunto de notícias falsas de destaque e estimou que o americano médio encontrou entre uma e três delas no mês anterior à eleição; a estimativa é provavelmente con-

servadora, porque o estudo acompanhou apenas 156 notícias. Ressalte-se que quantificar as contas que receberam ou replicaram uma postagem não revela quantas pessoas a leram, tampouco se esse estímulo exerceu qualquer efeito sobre suas ações.

Em síntese, como formulam Lazer et al. (2018): exposição não equivale a impacto, e a avaliação, a médio e longo prazo, do efeito da exposição a notícias falsas sobre o comportamento político é essencialmente inexistente na literatura. Mosleh e Rand (2022) lembram uma distinção parecida do lado de quem compartilha: exposição e compartilhamento são construtos fundamentalmente diferentes, a maior parte das pessoas compartilha só uma pequena fração do que vê, e olhar o que alguém compartilha dá um retrato muito limitado de seu ambiente de informação.

1.3 Elites políticas como fonte

A terceira pergunta sobre prevalência — de onde a informação falsa provém — desloca o foco do público em geral para as elites políticas. Mosleh e Rand (2022) argumentam que, depois da atenção às notícias de editores obscuros, uma forma diferente de desinformação passou a chamar a atenção: campanhas coordenadas por figuras e organizações públicas, que chamam de elites. Para estudá-la, construíram dois escores. O primeiro é um escore de falsidade para cada uma de 816 elites, calculado a partir das checagens profissionais do PolitiFact sobre a veracidade das declarações delas. O segundo é um escore de exposição para cada usuário do Twitter, calculado a partir dos escores de falsidade das elites que ele segue. Depois compararam a exposição com o que o próprio usuário compartilha. A qualidade das notícias compartilhadas foi medida pela confiabilidade do domínio, avaliada de duas formas: pela média de oito checadores profissionais e por uma amostra de 970 americanos escolhidos por cotas, com as notas de simpatizantes dos dois partidos balanceadas. A ideologia do usuário foi estimada pela ideologia das contas que ele segue. Quanto maior a exposição à desinformação de elites, menor a qualidade das notícias que o próprio usuário compartilha, mais à direita se situa sua inclinação política estimada. É uma correlação entre medidas, e os próprios autores dizem que a abordagem se baseia na rede imediata de contas seguidas e não capta a exposição além delas, como por retuítes de conexões ou por recomendações algorítmicas.

Lasser et al. (2022) analisam diretamente as postagens emitidas por parlamentares. Os autores reuniram os tuítes, sem retuítes, de membros do Congresso americano, do parlamento alemão e do parlamento britânico entre 2016 e março de 2022, extraíram os links e deram a cada um a nota do NewsGuard

para o domínio, de 0 (muito ruim) a 100 (jornalismo de qualidade excepcional); abaixo de 60, o domínio é considerado “não confiável”. Os republicanos compartilharam 9,1 vezes mais links para domínios não confiáveis que os democratas: 3,86% dos links, contra 0,43%. Na Alemanha, a CDU/CSU (democratas-cristãos) compartilhou 0,18% e o SPD (social-democratas) 0,03%; no Reino Unido, os conservadores 0,25% e os trabalhistas 0,05%, e os autores notam que esses números se baseiam em contagens muito baixas. A proporção de links não confiáveis dos republicanos passa de 2,4% em 2016–2018 para 5,5% em 2020–2022, enquanto a dos democratas segue em 0,4%; o resultado se repete com uma segunda base de confiabilidade de domínios, compilada independentemente do NewsGuard. A divergência entre os partidos parece peculiar aos Estados Unidos: não aparece na Alemanha nem no Reino Unido.

O que os autores propõem como explicação é uma possibilidade debatida: a motivação partidária, já que conteúdo depreciativo sobre o grupo político adversário aumenta as intenções de compartilhar entre partidários. Os autores sugerem também que o padrão entre os parlamentares pode contribuir para a assimetria que se observa no público, de duas maneiras: ao fornecer desinformação a partidários e ao legitimar o compartilhamento de informação não confiável. O NewsGuard só tem nota para parte dos links, e o estudo registra a cobertura: 46,5% dos links do Congresso americano, 58,8% do parlamento alemão e 39,2% do britânico, depois de excluídas as redes sociais e os buscadores. O que fica sem nota não é medido.

1.4 Posicionamento político e propensão à informação falsa: o que os estudos correlacionais mostram

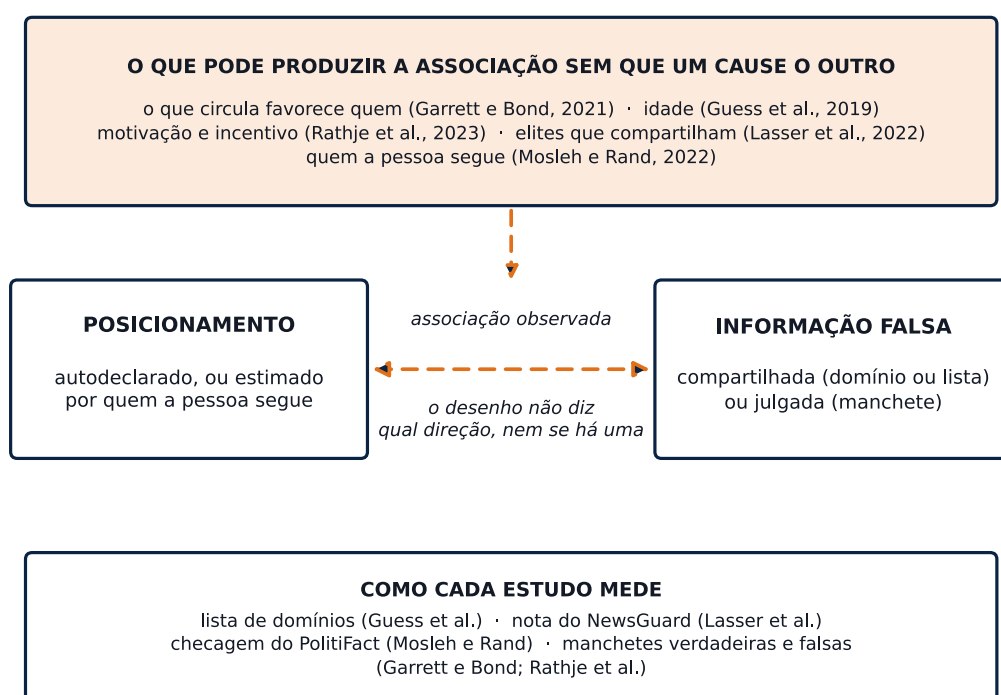
A pergunta recorrente na literatura é se há assimetria partidária: se quem se posiciona em determinado campo compartilha ou acredita mais em informação falsa. Os estudos verificados respondem com associações, e a Tabela 1.2 as reúne com o que cada desenho não permite concluir. A Figura 1.1 ilustra os limites dessa inferência.

Garrett e Bond (2021) adotaram um cuidado metodológico incomum na área. A evidência de que conservadores têm mais crenças falsas se apoiava em um conjunto relativamente estreito de temas, em muitos casos escolhidos de propósito para refletir falsidades promovidas por elites políticas conservadoras. Os autores, em vez disso, deixaram a escolha das afirmações a um critério externo: a cada duas semanas, um serviço identificou as 20 histórias políticas mais virais, 10 verdadeiras e 10 falsas, e um painel de 12 ondas, com grande amostra representativa de americanos, avaliou na semana seguinte afirmações curtas sobre essas histórias, ao longo de seis meses. Com as respostas,

calcularam dois índices, pela teoria de detecção de sinais: a sensibilidade, que é a capacidade de distinguir afirmações verdadeiras de falsas, e o viés de resposta, que é a propensão a rotular todas as afirmações como verdadeiras ou como falsas. Em dados de pesquisa nos Estados Unidos, os conservadores distinguiram verdades de falsidades pior que os liberais, e os autores atribuem parte da diferença ao ambiente informacional: as afirmações verdadeiras de alto perfil tendem a promover temas e candidatos favorecidos por liberais, e as narrativas falsas tendiam a favorecer causas conservadoras, de modo que, mesmo com igual viés, os liberais teriam desempenho melhor nesse ambiente.

FIGURA 1.1

O que uma associação entre posicionamento político e informação falsa mostra, e o que pode produzi-la sem que um cause o outro



Esquema original, sem dados. Tracejada: relação que os desenhos correlacionais não separam.

Nota. Esquema original do autor, a partir de Garrett e Bond (2021), Guess et al. (2019), Rathje et al. (2023), Lasser et al. (2022) e Mosleh e Rand (2022). Não representa dados. Tracejada: relação que os desenhos correlacionais não separam.

Rathje et al. (2023) acrescentam experimentos em que se pôde variar a consequência de acertar. Em quatro experimentos, com 3.364 participantes dos Estados Unidos, as pessoas liam manchetes políticas verdadeiras e falsas e respondiam se a afirmação da manchete era precisa; metade, sorteada, foi informada de que receberia um bônus de até um dólar conforme as respostas corretas, e a outra metade respondeu às mesmas perguntas sem incentivo. Os au-

tores chamam de viés partidário a diferença entre o número de manchetes favoráveis ao próprio partido e o de manchetes contrárias ao próprio partido que o participante julgou verdadeiras. Os incentivos financeiros por acurácia reduziram o viés partidário no julgamento de manchetes em cerca de 30%. Os conservadores foram menos acurados que os liberais em distinguir manchetes verdadeiras de falsas, mas os incentivos reduziram a diferença em 52%; e incentivar os participantes a identificar notícias que seriam apreciadas por seus aliados políticos diminuiu a acurácia. Parte do que aparece como diferença de crença, portanto, parece depender do que se reforça: acertar ou agradar.

TABELA 1.2

Cinco estudos sobre posicionamento político e informação falsa: o que mediram e o que não permitem concluir

Estudo	Amostra e país	Medida de informação falsa	Medida de posicionamento	Resultado	O que não permite
Guess et al. (2019)	Respondentes de pesquisa representativa ligados a dados do Facebook; EUA, 2016	Links de domínios de notícias falsas, por lista	Autodeclarado (partido, ideologia)	Raro (8,5%); mais entre republicanos e entre os mais velhos; a idade é o achado robusto	Atribuir a diferença a maior propensão: pode refletir a oferta pró-Trump
Garrett e Bond (2021)	Pesquisa nos EUA, afirmações ao longo de 6 meses	Julgamento de afirmações verdadeiras e falsas	Autodeclarado	Conservadores discriminaram pior; parte atribuída ao ambiente informacional	Concluir que há viés maior: o resultado depende do conjunto de afirmações
Mosleh e Rand (2022)	816 elites e usuários do Twitter, EUA	Escore de falsidade por checagens do PolitiFact	Estimado por quem o usuário segue	Exposição a elites falsas correlaciona com pior qualidade de notícias e com ideologia conservadora	Causalidade; um verificador só; exposição além de quem se segue
Lasser et al. (2022)	Parlamentares dos EUA, da Alemanha e do Reino Unido, 2016-2022	Nota do NewsGuard do domínio do link	Partido	Republicanos compartilham mais, e mais com o tempo; Alemanha e Reino Unido, contagens baixas	Inferir a veracidade da fala a partir do domínio da fonte; cobertura restrita de 39% a 59% dos linkss

Estudo	Amostra e país	Medida de informação falsa	Medida de posicionamento	Resultado	O que não permite
Rathje et al. (2023)	Quatro experimentos, n = 3.364, EUA	Julgamento de manchetes verdadeiras e falsas	Autodeclarado	Incentivo reduziu o viés partidário em cerca de 30% e a diferença em 52%	Generalizar de manchetes e participantes online

Nota. Dados compilados a partir dos artigos originais citados. As restrições da última coluna explicitam o que os delineamentos metodológicos não conseguem isolar.

1.5 Limitações: o que os desenhos não permitem

Os estudos reunidos nas seções anteriores têm três limites em comum, e é por eles que os números do capítulo devem ser lidos. Dos cinco estudos da Tabela 1.2, quatro são sobre os Estados Unidos e o quinto compara com Alemanha e Reino Unido; os dados de três deles vêm de uma plataforma, o Facebook ou o Twitter. O primeiro limite é, portanto, de amostra e de contexto: os achados são do sistema bipartidário americano, em eleições específicas, e Lasser et al. (2022) dizem que a divergência que observam parece peculiar aos Estados Unidos. O segundo é de causa: nenhum desses desenhos separa o que a pessoa vê, o que lhe é oferecido e o que ela escolhe, e a Figura 1.1 reúne as variáveis que podem produzir a associação sem que posicionamento e informação falsa se causem. O terceiro é de medida, e merece um quadro, porque é o que se toma emprestado da análise do comportamento para ler todos os outros.

CONCEITO EM USO

Comportamento verbal, audiência e autorrelato

Comportamento verbal é o comportamento reforçado pela mediação de outras pessoas (Skinner, 1957); a audiência é o estímulo discriminativo em cuja presença esse comportamento é caracteristicamente reforçado. Autorrelato, isto é, responder a uma escala, a uma pergunta de pesquisa ou a uma entrevista, é comportamento verbal, e está sob controle de quem pergunta e de como a pergunta é feita. Para Guerin (1994), escalas e questões de pesquisa dependem de quem pergunta e da cultura verbal em que a pesquisa é feita, porque atitudes dependem sempre de consequências sociais. O exemplo clássico é o de Hildum e Brown (1956): num levantamento telefônico com 40 estudantes homens de Harvard, o entrevistador dizia “Good” depois de respostas a favor, ou contra, o programa de Educação Geral, e o “Good” deslocou as respostas na direção reforçada; só 8 de 20 sujeitos notaram o que o entrevistador fazia, e todos rejeitaram que suas respostas tivessem sido influenciadas por ele.

A consequência para este capítulo é direta. A Tabela 1.2 traz três linhas em que o dado é uma resposta a uma pesquisa ou a uma tarefa de julgamento (Garrett e Bond, Rathje et al. e, em parte, Guess et al., em que a ideologia é autodeclarada), e uma resposta a uma pesquisa é comportamento sob controle da situação em que foi dada. O resultado de Rathje et al. ilustra precisamente essa dinâmica: quando a consequência de acertar muda, o julgamento verbal se altera. Um segundo problema de medida é o da seleção dos estímulos. Cameoka e Moreira (2021) relatam três estudos com universitários que avaliaram fotos de homens brancos e de homens negros. No Estudo 1, 75 participantes avaliaram, numa escala de 1 (mau) a 10 (bom), 10 fotos de homens brancos e 10 de homens negros com expressões neutras; o Estudo 2 teve 40 participantes, com faces discretamente sorridentes, e o Estudo 3, 30. A diferença entre as médias dos dois grupos de fotos foi significativa nos Estudos 1 e 2 e não no 3, em que se usaram as fotos de brancos com as avaliações mais altas e as de negros com as mais baixas nos dois primeiros estudos. Os autores tomam o Estudo 3 como demonstração do potencial de distorção de uma seleção de estímulos pouco rigorosa e sugerem que a seleção dos estímulos pode enviesar a mensuração. O que se mede depende do que se escolheu para ser medido.

Há ainda o limite do instrumento de checagem, que o Capítulo 2 trata por inteiro: a proporção de falsas entre as declarações checadas não é a proporção entre todas as declarações, e uma agência não é um sensor neutro. Mosleh e Rand (2022) usam um único verificador, o PolitiFact, e elites dos Estados Unidos. O leitor deve guardar, para os demais capítulos, o que esta seção estabelece: toda taxa de falsidade é função de quem foi medido, com que instrumento, em que situação, e nenhuma delas é, por si, a taxa de um campo político.

1.6 O que se sabe sobre o Brasil

O que as fontes verificadas dizem do Brasil é menos do que a seção 1.2 diz dos Estados Unidos, e de natureza diferente. Ricard e Medeiros (2020), num comentário, e não num estudo empírico, escrevem que uma retórica de negação dos riscos da COVID-19 foi endossada nos mais altos níveis do governo brasileiro e apoiada por desinformação; e registram que, segundo a Comissão Parlamentar de Inquérito (CPI da Pandemia), parte da desinformação sobre a pandemia atacava decisores e figuras públicas, como governadores, a mídia e especialistas em saúde, para deslegitimar quem apoiava o isolamento, e que a Comissão ainda não havia divulgado evidência de coordenação direta pelo chamado Gabinete do Ódio ou por membros do governo.

Carvalho et al. (2021) fizeram uma análise de conteúdo, com o programa IRa-MuTeQ, de postagens nos perfis de Trump e de Bolsonaro no Facebook e de pronunciamentos oficiais nos sites dos dois governos, desde o primeiro dia de notificação da Covid-19 em cada país. Os autores apontam similaridades temáticas e posicionamentos contrários às medidas defendidas pela OMS. Escrevem que não é possível afirmar categoricamente que houve alinhamento intencional entre os dois, mas que a análise mostra convergência nas estratégias de desinformação, e que a minimização dos riscos de contágio, sobretudo nas postagens do Facebook, é uma das marcas da similaridade. É uma análise de dois emissores, e não um dado de prevalência.

Recuero et al. (2025) analisaram 811 checagens de três agências brasileiras sobre as eleições de 2018 e 2022, por análise de conteúdo e métodos mistos, comparando temáticas, tipos de desinformação e as fontes que legitimavam o conteúdo. Apontam uma transição de conteúdos totalmente falsos para enganosos, e de conteúdos sem fontes para conteúdos que usam fontes reais; leem a mudança das plataformas dos conteúdos checados, do Facebook para o WhatsApp e o Twitter entre 2018 e 2022, como adaptação da desinformação a espaços mais fechados, em que é mais difícil circular discursos. Registram, igualmente, que a literatura revisada aponta maior volume de desinformação gerada por apoiadores de Bolsonaro em ambos os pleitos, ainda que ambas as campanhas tenham figurado concomitantemente como polos emissores e alvos de desinformação. A Lupa (Agência Lupa, 2026) descreve o mesmo deslocamento em sua própria base: em 2018, o Facebook aparecia em 76,5% das verificações; em 2022, o WhatsApp estava presente em 92,4%.

Dois limites separam essas fontes do que a seção 1.2 mostra para os Estados Unidos. Nenhuma delas é um estudo de compartilhamento com registro de comportamento, como o de Guess et al. (2019), nem uma comparação de partidos pelo que seus parlamentares compartilham, como a de Lasser et al.: o que há são checagens, o que é curado por uma agência, e uma análise de conteúdo de dois emissores. E a regra do Capítulo 2, de que a proporção de falsas entre as declarações checadas não é a proporção entre todas as declarações, vale aqui integralmente. Que a desinformação brasileira tenha o formato da americana, ou o contrário, é hipótese, e este capítulo não a assume.

1.7 Discurso de ódio: definições, medidas e dados

Entre as fontes verificadas, a expressão 'discurso de ódio' têm uma definição explícita em Hangartner et al. (2021): linguagem hostil contra um grupo ou pessoa por característica inata, real ou percebida; os autores selecionaram tu-

ítes xenófobos e racistas por combinação de dicionário, análise de sentimento e anotação manual. Os demais estudos operacionalizam de outro modo, e a Tabela 1.3 reúne o que cada um mediu.

TABELA 1.3

Como cinco estudos verificados medem discurso de ódio, e o que cada medida não permite

Estudo	Medida de “discurso de ódio”	Amostra	Resultado	O que a medida não permite
Siegel et al. (2021)	Dicionário aumentado por aprendizado de máquina e classificação por comunidades do Reddit; o dicionário exige insulto ou termo depreciativo	Mais de 750 milhões de tweets sobre a eleição americana de 2016 e quase 400 milhões de tweets de uma amostra aleatória de usuários americanos	Sem aumento persistente na campanha nem depois da eleição	Falar de outras plataformas; a definição é conservadora
Marino-Fages e Martínez-Pozo (2026, documento de trabalho)	Classificador BERTimbau e dicionário, por alvo (homofobia, racismo, sexismo, insultos, ódio político)	Mais de 37,6 milhões de tweets geolocalizados no Brasil, 2017-2019	Aumento depois da eleição de 2018, maior onde Bolsonaro teve menos votos, em homofobia e sexismo	Medir o custo social percebido; a acurácia do classificador é de 77% a 79%
Álvarez-Benjumea e Winter (2018)	Nota de hostilidade (7 ou mais) dada a comentários de fórum experimental	180 participantes de plataforma de crowdsourcing	Comentários hostis: 5,88% na base, 0,99% com remoção moderada; anônimo contra-fala sem efeito significativo	Generalizar: participantes pagos, fórum anônimo
Hangartner et al. (2021)	Tuítes xenófobos e racistas, selecionados por dicionário, sentimento e anotação	1.350 usuários do Twitter	Contra-fala baseada em empatia aumentou a exclusão e reduziu a criação de discurso xenófobo	Efeito sobre quem apenas lê; humor e aaviso de consequências sem efeito consistente
Fazzano e Gallo (2015)	Homofobia como classe de respostas (bater, insultar, torturar, espancar, excluir, discriminar), definida pela função	Dados do Disque 100 de 2011: 6.809 relatos de violações contra LGBT	Registros de denúncias oficiais; não mensuram a prevalência direta da agressão	Prevalência: são relatos, e não o fenômeno

Nota. Números dos artigos. Marino-Fages e Martínez-Pozo é documento de trabalho que, segundo a capa, não passou por revisão de pares. Os estudos são tratados no Capítulo 5 (seção 5.5), no Capítulo 10 e no Capítulo 12.

Quatro pontos organizam a tabela. O primeiro é que a definição decide o número: Siegel et al. dizem que sua definição de dicionário é conservadora, pois exige que o tuíte contenha insulto ou termo depreciativo, e Marino-Fages e Martínez-Pozo classificam por alvo e obtêm um aumento concentrado em homofobia e sexismo, sem padrão comparável para insultos gerais, ódio político ou alvos políticos. O segundo é a plataforma e o recorte da amostra: o estudo americano se limita ao Twitter e adverte que observar certo discurso num momento não significa que ele tenha aumentado, dado o caráter intermitente da plataforma, em que tópicos ganham destaque por pouco tempo e voltam ao equilíbrio; o brasileiro usa só tweets publicados às segundas-feiras e geolocalizados no Brasil, e assume a premissa de que esses dois recortes representam o universo da plataforma. O terceiro é que “discurso” e “ato” são medidas diferentes, e nenhum dos cinco estudos mede ato contra o grupo (a lacuna é discutida na seção 5.5). O quarto é que os resultados não se contradizem nem se confirmam: medem coisas diferentes em populações diferentes.

Do lado da análise do comportamento, a literatura é escassa. Mizael (2018) encontrou apenas quatro artigos analítico-comportamentais sobre homossexualidade em 13 periódicos e aponta como lacuna as análises funcionais de comportamentos preconceituosos, como piadas mantidas pela comunidade verbal; e Fazzano e Gallo (2015) definem a homofobia como classe de respostas de topografias diversas, definida pela função, e não pela forma. É a definição de que o Capítulo 5 parte.

1.8 Interpretações rivais e limites

A leitura proposta neste capítulo é que os dados empíricos revelam magnitudes inferiores ao que o debate público supõe, e que cada cifra é função estrita do instrumento utilizado. Essa perspectiva confronta três interpretações rivais. A primeira sustenta que a assimetria partidária seria um fato consolidado e que o capítulo adotaria um ceticismo desmedido; o teste decisivo entre essas leituras exigiria um delineamento que mantivesse sob controle rigoroso a oferta total de conteúdo e a exposição individual efetiva — ressalva que os próprios autores de Garrett e Bond (2021) e Guess et al. (2019) formulam. A segunda tese rival argumenta que, sem checagem profissional de fatos, qualquer medição é inviável, tornando as agências o único termômetro disponível; a resposta do Capítulo 2 é que a checagem mede o que foi checado, e não o que circula. A terceira é que a definição pela intenção torna a medida impossível e deveria ser abandonada; a de domínio, que os estudos usam, é uma resposta pragmática, e aqui é tomada como tal.

Os limites do capítulo são os das fontes. A maioria das medidas é de Estados Unidos e de uma plataforma, no período de uma ou duas eleições; a única base brasileira de grande escala, entre as de discurso, é um documento de trabalho sem revisão de pares; e os números de prevalência de informação falsa são, no melhor dos casos, de exposição e de compartilhamento, e não de efeito. Não se compara o Brasil com os Estados Unidos, e o compartilhamento de uma fonte não é tratado como crença nem como ato.

Síntese

"Cada definição de informação falsa impõe exigências metodológicas próprias (intenção do emissor, origem do conteúdo, confiabilidade do domínio ou veracidade factual), e as pesquisas tendem a substituir a análise da declaração pela classificação da fonte em virtude da inacessibilidade da intenção privada (1.1). A circulação empírica de notícias falsas por compartilhamento individual revelou-se rara nos dados de Guess et al., com maior incidência entre idosos, e exposição passiva não equivale a impacto causal (1.2). A assimetria partidária registrada entre parlamentares nos Estados Unidos pelo critério de compartilhamento de domínios parece constituir fenômeno peculiar àquele contexto institucional (1.3). As relações entre posicionamento ideológico e desinformação têm caráter correlacional, dependem do conjunto de itens selecionados e do ambiente de oferta, sendo que parte substancial da disparidade de acurácia se altera quando as contingências de reforço são manipuladas (1.4). autorrelato em pesquisas consiste em comportamento verbal sob controle das contingências da coleta, e a amostragem arbitrária de estímulos pode enviesar o resultado das medidas (1.5). No Brasil, o quadro empírico limita-se a análises de conteúdo, relatórios de conjuntura e bases de agências checadoras, inexistindo registros comportamentais diretos de compartilhamento em rede aberta (1.6). Por fim, as mensurações de discurso de ódio variam radicalmente conforme os critérios lexicais e a plataforma de coleta adotados, não capturando a emissão de atos concretos contra os grupos-alvo (1.7).

Os capítulos subsequentes mobilizam diretamente este arcabouço: os Capítulos 2 e 3 operam sob o princípio de que toda taxa de falsidade é função do instrumento de medida; o Capítulo 5 adota as definições funcionais da seção 1.7; o Capítulo 6 fundamenta-se na evidência de que o compartilhamento representa uma fração ínfima da exposição; e o Capítulo 12 examina a discrepância entre o que o indivíduo relata verbalmente e o que ele efetivamente faz diante de contingências reais.

Referências

- Agência Lupa. (2026, 27 de julho). *Anatomia da desinformação eleitoral* [Relatório]. https://uploads.agencialupa.org/2026/07/RELATORIO-ANATOMIA-DA-DESINFORMACAO-ELEITORAL_LUPA-FINAL.pdf
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211–236. <https://doi.org/10.1257/jep.31.2.211>
- Álvarez-Benjumea, A., & Winter, F. (2018). Normative change and culture of hate: An experiment in online environments. *European Sociological Review*, 34(3), 223–237. <https://doi.org/10.1093/esr/jcy005>
- Cameoka, M. C., & Moreira, M. B. (2021). Preconceito racial: viés na mensuração de atitudes produzido por controle de estímulos. *Acta Comportamentalia*, 29(1), 93–112. <https://doi.org/10.32870/ac.v29i1.78782>
- Carvalho, P. R., Sousa, P. C. C., & Schneider, M. A. F. (2021). Desinformação na pandemia de Covid-19: similitudes informacionais entre Trump e Bolsonaro. *Em Questão*, 27(3), 15–41. <https://doi.org/10.19132/1808-5245273.15-41>
- Fazzano, L. H., & Gallo, A. E. (2015). Uma análise da homofobia sob a perspectiva da análise do comportamento. *Temas em Psicologia*, 23(3), 535–545. <https://doi.org/10.9788/tp2015.3-02>
- Garrett, R. K., & Bond, R. M. (2021). Conservatives' susceptibility to political misperceptions. *Science Advances*, 7(23), Article eabf1234. <https://doi.org/10.1126/sciadv.abf1234>
- Guerin, B. (1994). Attitudes and beliefs as verbal behavior. *The Behavior Analyst*, 17(1), 155–163. <https://doi.org/10.1007/bf03392661>
- Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, 5(1), Article eaau4586. <https://doi.org/10.1126/sciadv.aau4586>
- Hangartner, D., Gennaro, G., Alasiri, S., Bahrich, N., Bornhoft, A., Boucher, J., Demirci, B., Derksen, L., Hall, A., Jochum, M., Munoz, M. M., Richter, M., Vogel, F., Wittwer, S., Wüthrich, F., Gilardi, F., & Donnay, K. (2021). Empathy-based counterspeech can reduce racist hate speech in a social media field experiment. *Proceedings of the National Academy of Sciences*, 118(50), Article e2116310118. <https://doi.org/10.1073/pnas.2116310118>
- Hildum, D. C., & Brown, R. W. (1956). Verbal reinforcement and interviewer bias. *The Journal of Abnormal and Social Psychology*, 53(1), 108–111. <https://doi.org/10.1037/h0040518>
- Lasser, J., Aroyehun, S. T., Simchon, A., Carrella, F., Garcia, D., & Lewandowsky, S. (2022). Social media sharing of low-quality news sources by political elites. *PNAS Nexus*, 1(4), Article pgac186. <https://doi.org/10.1093/pnasnexus/pgac186>
- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094–1096. <https://doi.org/10.1126/science.aao2998>

- Marino-Fages, D., & Martínez-Pozo, A. (2026). *Hate in the Tropics: Political leaders and the social acceptability of online hate speech* (Documento de trabajo N.º 391). RedNIE. <https://rednie.eco.unc.edu.ar/files/DT/391.pdf>
- Mizael, T. M. (2018). Perspectivas analítico-comportamentais sobre a homossexualidade: análise da produção científica. *Revista Perspectivas*, 9(1), 15–28. <https://doi.org/10.18761/PAC.2017.011>
- Mosleh, M., & Rand, D. G. (2022). Measuring exposure to misinformation from political elites on Twitter. *Nature Communications*, 13(1), Article 7144. <https://doi.org/10.1038/s41467-022-34769-6>
- Rathje, S., Roozenbeek, J., Van Bavel, J. J., & van der Linden, S. (2023). Accuracy and social motivations shape judgements of (mis)information. *Nature Human Behaviour*, 7(6), 892–903. <https://doi.org/10.1038/s41562-023-01540-w>
- Recuero, R., Tavares, C., Volcan, T., Pozzebon, M., & Dutra, M. (2025). A desinformação narrada pela checagem: estudo das eleições de 2018 e 2022. *E-Compós*, 28, 1–26. <https://doi.org/10.30962/ecomps.3039>
- Ricard, J., & Medeiros, J. (2020). Using misinformation as a political weapon: COVID-19 and Bolsonaro in Brazil. *Harvard Kennedy School Misinformation Review*, 1(2), Article 013. <https://doi.org/10.37016/mr-2020-013>
- Siegel, A. A., Nikitin, E., Barberá, P., Sterling, J., Pullen, B., Bonneau, R., Nagler, J., & Tucker, J. A. (2021). Trumping hate on Twitter? Online hate speech in the 2016 U.S. election campaign and its aftermath. *Quarterly Journal of Political Science*, 16(1), 71–104. <https://doi.org/10.1561/100.00019045>
- Skinner, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts.
- Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policymaking (Council of Europe Report DGI[2017]09). Council of Europe. <https://rm.coe.int/information-disorder-toward-an-interdisciplinary-framework-for-research/168076277c>



ENCICLOPÉDIA IW4

Os conceitos deste livro, um verbete de cada vez

Reforçamento, controle de estímulos, equivalência, prática baseada em evidências: verbetes técnicos escritos com clareza, com as fontes ao pé de cada um. Para estudar, revisar e citar.

80+ verbetes abertos

Acesse, de graça e sem cadastro

walden4.com.br/enciclopedia



2

O que dizem as agências de checagem (e o que elas não podem dizer)

Em julho de 2026, a Agência Lupa divulgou o relatório *Anatomia da desinformação eleitoral*, com a base de suas checagens em cinco ciclos eleitorais, de 2016 a 2024. Entre os 23 políticos com mais de 30 frases avaliadas, Jair Bolsonaro teve a maior proporção de frases classificadas como “Falso”: 102 de 236, ou 43,2%, nas campanhas presidenciais de 2018 e 2022. Luiz Inácio Lula da Silva aparece em sétimo lugar: 22 de 83 frases, ou 26,5%, em campanhas eleitorais que disputou “recentemente” (Agência Lupa, 2026, p. 12, 19–20).

Em outra frente, o Aos Fatos manteve, de 2019 a 2022, um contador das declarações “falsas ou distorcidas” de Bolsonaro. Em 22 de dezembro de 2022, informou que ele deixava a presidência com 6.675 (Ribeiro & Menezes, 2022); em julho de 2025, o mesmo veículo falou em 6.685 (Menezes & Faustino, 2025). Para Lula, o Aos Fatos não tem contador equivalente: publicou, no fim de 2023 e de 2024, levantamentos das alegações que ele mais repetiu (Ribeiro, 2023; Menezes, 2024).

Colocar 43,2% ao lado de 26,5%, ou 6.685 ao lado de um ranking de alegações repetidas, dá a impressão de comparar. Sustenta-se neste capítulo que, com os dados publicados hoje, nenhuma das duas operações constitui uma comparação legítima, explicitando-se o que faltaria para que fosse."

Não é ceticismo contra as agências. É a leitura que elas próprias pedem: a Lupa escreve, no mesmo relatório, que optou por não produzir rankings de quem “mentiu mais”, porque o volume de checagens sobre cada político depende da exposição pública, da frequência de entrevistas e debates, do cargo e do estilo de comunicação (Agência Lupa, 2026).

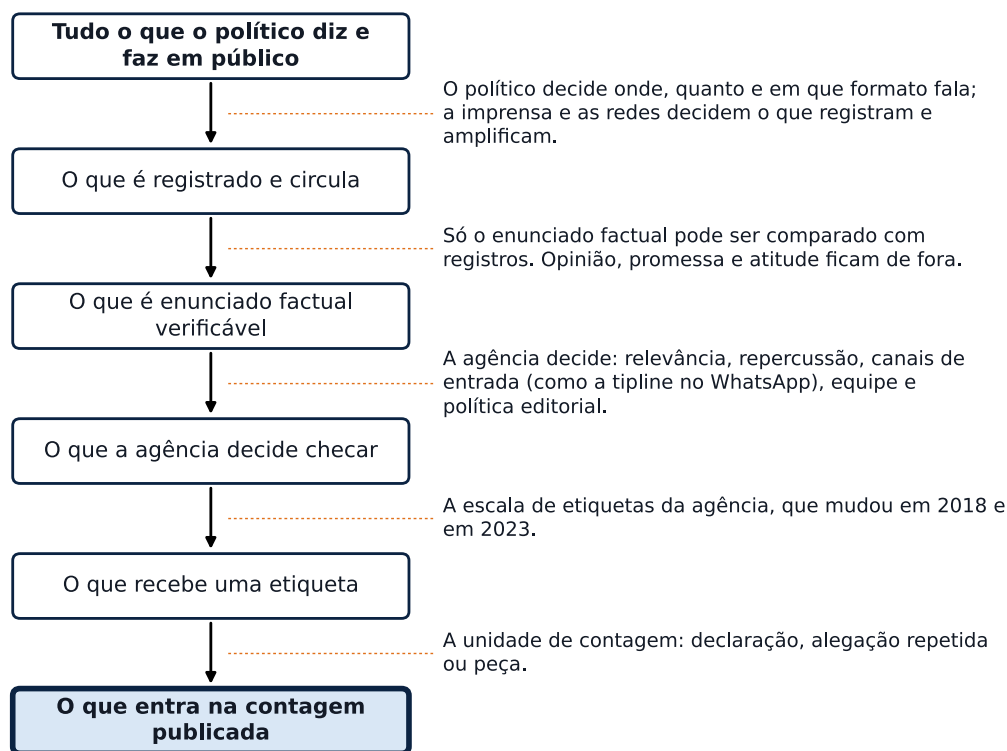
O capítulo cumpre também uma função no livro: fixa as regras com que os capítulos seguintes leem dados de agências. Quando uma fala aparecer como “classificada como falsa pela agência X”, quer dizer exatamente isso, e nada além.

2.1 Duas séries de contingências

Uma fala de político que chega à contagem de uma agência passou por duas séries de contingências. A primeira é a do político: onde, para quem, com que frequência e em que formato ele fala. A segunda é a da agência: o que entra no seu radar, o que ela considera relevante e verificável, que escala usa para classificar e como conta. A Figura 2.1 resume o percurso. Cada seta é uma seleção, e só o último estágio chega ao leitor.

FIGURA 2.1

Do enunciado à classificação: onde a seleção acontece



Só o último estágio chega ao leitor: as proporções publicadas são frações do que passou pelos anteriores.

Nota. Esquema original do autor. Não representa proporções: os estágios não têm tamanho conhecido. Os mecanismos anotados nas setas vêm da própria Lupa (Agência Lupa, 2026) e, para a unidade de contagem, das fontes da Tabela 2.1.

Três pontos da figura merecem detalhe, porque as próprias agências os documentam.

Primeiro, o que é checado depende de quanto o político se expõe. A Lupa observa que candidatos que falam mais e são mais incisivos produzem mais afirmações passíveis de verificação e, por isso, podem “ter mais ‘Falsos’ e despontar erroneamente como o mais mentiroso” (Agência Lupa, 2026, p. 19). O contador do Aos Fatos mostra o mesmo efeito na escala de um dia: uma única entrevista, ao Flow Podcast em 8 de agosto de 2022, rendeu ao menos 26 declarações falsas ou distorcidas em cinco horas e 21 minutos, e o recorde do mandato foi 15 de janeiro de 2021, com 61 falas checadas (Santos, 2022). Os dois casos sugerem que a contagem acompanha, em parte, o tempo de exposição.

Segundo, o que entra no radar da agência depende de seus canais de entrada. A Lupa mantém uma tipline no WhatsApp e integra programas de verificação da Meta, o que, nas suas palavras, tende a concentrar parte do fluxo de conteúdos recebidos em plataformas da Meta, de modo que os resultados sobre plataformas devem ser lidos como “um retrato do universo monitorado pela Lupa” (Agência Lupa, 2026, p. 82). E a base do relatório reflete apenas o que a agência publicou na editoria de Eleições, condicionado a escolhas editoriais e a critérios de relevância jornalística.

Terceiro, a seleção pode ser lida como comportamento da própria agência. Kniess et al. (2024) partem de um cenário em que o jornalismo enfrenta uma crise de legitimidade e de confiança e perguntam em que medida duas agências ligadas a empresas de comunicação tradicionais, o Estadão Verifica e o Fato ou Fake, buscam reforçar a própria autoridade epistêmica, isto é, a credibilidade como fonte do que é verdadeiro, por meio da checagem. Para responder, coletaram todos os textos que as duas agências publicaram com o nome de Bolsonaro no título entre janeiro e maio de 2022 e com o de Lula entre janeiro e maio de 2023, 119 checagens ao todo (95 do Estadão Verifica e 24 do Fato ou Fake), e submeteram esse corpus à análise de conteúdo. As agências concentraram a cobertura em assuntos semelhantes nos dois governos, e a apuração deu ênfase a fontes do próprio jornalismo, isto é, a materiais próprios ou de outras empresas de comunicação tradicionais. Segundo os autores, isso sugere limites à transparência da apuração, porque muitas das informações usadas já passaram por filtros editoriais nem sempre claros ao público.

Em linguagem analítico-comportamental, a checagem é comportamento verbal do chegador. Está sob controle de estímulos antecedentes (o enunciado do político e a audiência), mas também de contingências de reforçamento institucionais (como a política editorial) e de consequências que extrapolam o enunciado checado. Isso não a torna arbitrária. Torna a contagem dependente de um processo de seleção que precisa ser descrito para ser lido. Que entre

essas consequências esteja, por exemplo, a credibilidade perante a audiência é uma hipótese compatível com Kniess et al. (2024), não um dado deste capítulo.

2.2 O que cada fonte conta

Antes de comparar números, pergunte de cada um o que foi contado. O Quadro 2.1 reúne as cinco perguntas e a Tabela 2.1 as responde para as fontes usadas neste capítulo.

QUADRO 2.1

Cinco perguntas antes de comparar contagens

1. Qual é a *unidade*: frase, declaração, alegação repetida, peça ou menção?
2. Qual é o *denominador*, isto é, o que foi checado, e ele é publicado?
3. Qual é o *período* e a *situação*: campanha, mandato, ano civil?
4. Qual é a *escala* de classificação, e ela mudou?
5. *Quem escolheu* o que entrou, e por qual regra?

TABELA 2.1

O que cada fonte conta e o que publica como denominador

Fonte	Unidade contada	O que foi publicado	Denominador publicado	Período e situação
Lupa, relatório (Agência Lupa, 2026)	Frase checada (unidade de análise dentro de uma checagem)	Frases “Falso” por político, entre os que têm mais de 30 frases	Sim, por político: 236 (Bolsonaro) e 83 (Lula)	Campanhas eleitorais de 2016 a 2024; Bolsonaro, 2018 e 2022; Lula, ciclos não discriminados
Aos Fatos, contador de Bolsonaro (Santos, 2022; Ribeiro & Menezes, 2022; Menezes & Faustino, 2025)	Declaração “falsa ou distorcida” (“imprecisa” e “incorreta”, em 2025)	6.042 (ago./2022); Não 6.675 (dez./2022); 6.685 (jul./2025)		Mandato, jan./2019 a dez./2022
Aos Fatos, levantamentos de Lula (Ribeiro, 2023; Menezes, 2024)	Alegação repetida	Ranking das mais repetidas, com o número de ocasiões (21 vezes em 2023; ao menos 24 em 2024)	Não	Ano civil de 2023 e de 2024; mandato em curso

Fonte	Unidade contada	O que foi publicado	Denominador publicado	Período e situação
Aos Fatos, balanço de dez anos (Menezes & Faustino, 2025)	Checagens publicadas; menções em peças de desinformação	Mais de 19 mil checagens; cerca de 15 mil declarações de políticos e autoridades; menções: Lula, 677, e Bolsonaro, 566	Total geral, sem recorte por político nem por mandato	2015 a 2025
Kniess et al. (2024)	Checagem (peça) com o nome do presidente no título	119 checagens: 95 do Estadão Verifica e 24 do Fato ou Fake	Sim, do corpus	Jan.-mai./2022 (Bolsonaro) e jan.-mai./2023 (Lula)
Recuero et al. (2025)	Checagem (peça)	811 checagens de três agências	Sim, do corpus	Eleições de 2018 e 2022

Nota. Todas as fontes são checagens publicadas pelas próprias agências ou estudos sobre elas; em todas, quem escolhe o que entra é a agência (pergunta 5).

A tabela torna evidente o que o debate público costuma ocultar. Apenas a Lupa publica numerador e denominador individualizados por político — res- tritos, contudo, a períodos de campanha, e não a mandatos. O Aos Fatos, por sua vez, divulga para Bolsonaro um numerador desprovido de denominador; e, para Lula, reporta alegações repetidas, que constituem uma unidade de aná- lise distinta. Seus totais gerais (mais de 19 mil checagens; cerca de 15 mil de- clarações de políticos e autoridades e quase 4.000 peças de desinformação des- de 2015) não vêm recortados por político nem por mandato (Menezes & Faus- tino, 2025) e não servem de denominador para nenhum dos dois.

2.3 Por que a proporção entre as checadas não é a proporção entre as ditas

A fração publicada por uma agência é uma proporção entre enunciados que passaram pela seleção da Figura 2.1. Seja p a proporção de enunciados factu- ais falsos no que o político diz; sF , a probabilidade de que um enunciado falso seja checado; e sV , a de que um verdadeiro seja. A proporção observada entre as checadas é

$$p \cdot sF \div (p \cdot sF + (1 - p) \cdot sV),$$

que só coincide com p quando $sF = sV$ — ou seja, na hipótese de a agência che- car enunciados falsos e verdadeiros com idêntica probabilidade. Nenhum dos relatórios publicados autoriza essa premissa de igualdade, tampouco permite estimar a razão de seleção $r = sF \div sV$. A Lupa informa que checa também para confirmar o que é verdadeiro (“uma função central (e frequente) do fact-

checking”) e que 34,4% das frases checadas receberam “Verdadeiro” (Agência Lupa, 2026, p. 12, 29), o que mostra que a seleção não é só de suspeitas, mas não diz com que probabilidade relativa.

A Tabela 2.2 mostra, com números hipotéticos, o que a razão r faz com a proporção observada.

TABELA 2.2

Ilustração hipotética: a mesma fórmula com razões de seleção diferentes

Taxa real de falsidade (p)	Razão de seleção ($r = sF \div sV$)	Proporção observada entre as checadas
20%	1,0	20,0%
20%	1,5	27,3%
20%	2,0	33,3%
20%	3,0	42,9%
10%	4,0	30,8%

Nota. Valores hipotéticos, escolhidos para ilustrar a fórmula; não são dados nem estimativas sobre nenhum político. A proporção observada é $p \cdot r \div (p \cdot r + 1 - p)$.

A última linha é o ponto. Um político com metade da taxa real de falsidade (10% contra 20%) aparece com proporção observada maior (30,8%) do que outro com taxa real de 20% e razão 1,5 (27,3%), se a seleção favorecer mais a checagem do que é falso no primeiro. O mesmo raciocínio se aplica a discrepâncias entre agências distintas ou entre diferentes intervalos temporais. A conclusão não é a invalidade desses dados, mas sim que a diferença entre proporções observadas combina dois processos independentes — o repertório emitido pelo político e o filtro de seleção da agência —, variáveis que as bases publicadas não permitem isolar.

Há ainda uma segunda fonte de imprecisão. Frases de um mesmo debate, entrevista ou sabatina não são independentes: partilham tema, audiência e momento. A Lupa lembra que, nas checagens de debates e entrevistas, uma única reportagem pode analisar dezenas de frases de candidatos (Agência Lupa, 2026). O caso do Flow Podcast (seção 2.1) é a mesma coisa vista pelo lado do Aos Fatos.

2.4 O que a Lupa permite dizer

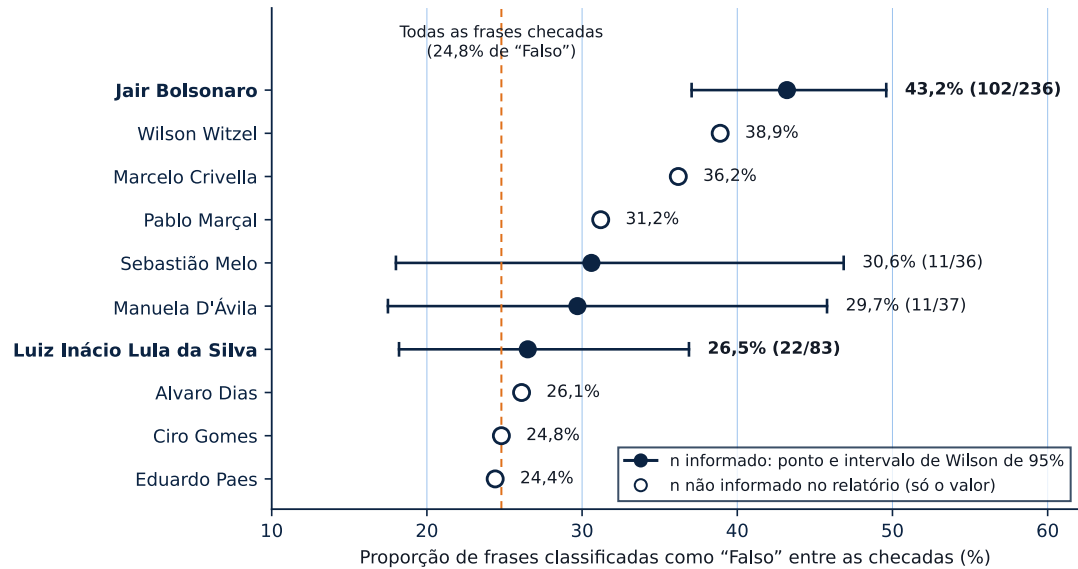
A Lupa é a única fonte examinada que publica numerador e denominador por político, com a mesma escala para todos. É, por isso, o material mais próximo de uma comparação. A Tabela 2.3 reúne o que o relatório informa sobre os dois presidentes, e a Figura 2.2 os situa entre os dez políticos do ranking publicado.

TABELA 2.3
O que o relatório da Lupa informa sobre Bolsonaro e Lula

	Jair Bolsonaro	Luiz Inácio Lula da Silva
Ciclos considerados	Campanhas presidenciais de 2018 e 2022	“Campanhas eleitorais que disputou recentemente”; ciclos não discriminados
Frases checadas	236	83
Frases “Falso”	102	22
Proporção de “Falso”	43,2%	26,5%
Posição entre os 23 políticos ^{1º} com mais de 30 frases		7º
Intervalo de Wilson de 95%, 37,1% a 49,6% calculado pelo autor		18,2% a 36,9%
Padrão das falsidades, segundo a análise linguística da Lupa	Números incorretos; afirmações absolutas e enganosas; “nunca”, “jamais”, “nenhum”, “sempre”	Reivindicação de autoria ou de pioneirismo de políticas públicas

Nota. Fonte: Agência Lupa (2026). O intervalo de Wilson delimita a margem de confiança de 95% para uma proporção sob a premissa teórica de que as frases checadas representassem uma amostra aleatória independente do repertório emitido pelo falante. Como os itens checados partilham contextos comuns e resultam de filtros editoriais deliberados (seção 2.3), essa condição de independência não é estritamente satisfeita; o intervalo opera aqui apenas como estimador da ordem de grandeza da imprecisão decorrente do tamanho da amostra, e não como teste de hipótese confirmatório.

FIGURA 2.2
Proporção de frases classificadas como “Falso” entre as checadas pela Lupa, por político



Nota. Valores do gráfico da p. 20 do relatório da Lupa, que considera apenas atores com mais de 30 frases checadas (Agência Lupa, 2026); o número de frases vem do texto e só está disponível para quatro políticos. A ordem é a do relatório e não mede “quem mente mais” (seção 2.3). Os políticos são de partidos, cargos e esferas diferentes. Intervalos de Wilson de 95% calculados pelo autor.

Quatro observações impedem que a Tabela 2.3 vire uma conclusão apressada.

É frase, não fala nem peça. A unidade do relatório é a frase checada dentro de uma checagem. A base tem 632 publicações, 397 checagens e 235 verificações, que geraram 3.366 unidades de análise, 3.080 de checagens e 286 de verificações (Agência Lupa, 2026).

É campanha, não mandato. Os 236 e os 83 vêm de campanhas eleitorais. Para Bolsonaro, os ciclos são os de 2018 e 2022, em que ele foi candidato e, no segundo, presidente candidato à reeleição. Para Lula, o relatório fala em campanhas que disputou “recentemente” e não os discrimina. A situação do falante (candidato, presidente, oposição) muda o que se diz e o que se checa, e a tabela não a controla.

“Falso” é uma etiqueta entre várias. No conjunto da base, “Falso” soma 765 registros (24,8%); “Exagerado”, 502 (16,3%); e “Verdadeiro, Mas... / Falta Contexto”, 434 (14,1%) (Agência Lupa, 2026). Os dois últimos, juntos, somam 30,4%, mais do que “Falso”, e a Lupa observa que, em quatro dos cinco ciclos, eles superaram “Falso” (p. 27). O relatório publicado não discrimina a distribuição das demais etiquetas individualizada por político. Uma comparação feita só com “Falso” deixa de fora a maior parte do que a agência classifica como problemático no conjunto.

A escala mudou. A Lupa começou com oito etiquetas; incorporou “Subestimado” em 2018; e, em 2023, “De olho”, “Ainda é cedo” e “Verdadeiro, Mas...” deram lugar a “Falta Contexto” (Agência Lupa, 2026, p. 28). Para as análises, os próprios pesquisadores da Lupa fundiram “Verdadeiro, Mas...” e “Falta Contexto”. Qualquer série que acesse 2023 precisa fazer o mesmo.

■ DADO	Na base da Lupa, 43,2% das 236 frases checadas de Bolsonaro nas campanhas presidenciais de 2018 e 2022 foram classificadas como “Falso”, contra 26,5% das 83 frases checadas de Lula em campanhas que disputou recentemente. A proporção de “Falso” na base inteira é de 24,8%.
■ INTERPRETAÇÃO	Entre as frases que a Lupa escolheu checar em campanhas, a proporção de “Falso” foi maior para Bolsonaro do que para Lula e do que a média da base. Os intervalos de Wilson (37,1% a 49,6% e 18,2% a 36,9%) quase se tocam; indicam imprecisão considerável, sobretudo para Lula.
■ HIPÓTESE	Parte da diferença pode refletir estilos de fala (a Lupa descreve números e afirmações absolutas em um caso e autoria e pioneirismo em outro), a situação do falante em cada ciclo e a razão de seleção <i>r</i> da seção 2.3. Nenhuma das três foi testada aqui.

2.5 O que o Aos Fatos permite dizer

O Aos Fatos publica dois tipos de contagem sobre os presidentes, e cada um tem unidade e critério próprios: um contador contínuo, mantido entre 2019 e 2022, e levantamentos anuais sobre Lula. As duas passagens abaixo dizem o que cada um permite concluir.

O contador. Entre 2019 e 2022, o Aos Fatos manteve um contador das alegações falsas do presidente, que monitorava “entrevistas, lives, discursos e postagens desde a posse, em janeiro de 2019” (Santos, 2022). Em 17 de agosto de 2022 marcava ao menos 6.042; em 22 de dezembro, ao fim do mandato, 6.675, “uma média de 4,6 por dia” (Ribeiro & Menezes, 2022); em julho de 2025, o texto do mesmo veículo fala em 6.685, ora como declarações “falsas ou imprecisas”, ora como “incorretas” (Menezes & Faustino, 2025). Nos textos publicados pelo veículo, não há justificativa explícita para a discrepância de dez registros nem a especificação do critério adotado para a inclusão do rótulo 'imprecisa'.

Duas coisas, porém, são claras. A primeira é que o contador publica um numerador sem denominador: não há, nos textos divulgados pelo veículo, o total de declarações do presidente que foram checadas, de modo que 6.675 não expressa a proporção de nenhum conjunto. A segunda é que a unidade, a declaração, é cumulativa e acompanha a exposição, como mostrou o caso do Flow Podcast. A média de 4,6 por dia é coerente com 1.461 dias de mandato ($6.675 \div 1.461 = 4,57$, cálculo do autor), mas é média por dia de mandato, e não por fala nem por hora de exposição.

Os levantamentos de Lula. Em 2023, o levantamento foi feito “com base nos discursos, entrevistas e transmissões ao vivo do presidente” (Ribeiro, 2023). Em 2024, o Aos Fatos “analisou todos os discursos, pronunciamentos e entrevistas” de Lula e separou as declarações passíveis de checagem; no decorrer do ano, checou as mais repetidas, e, no fim de dezembro, as classificadas como “Falso” ou “Não é bem assim” foram contadas para formar um ranking (Menezes, 2024).

A unidade, portanto, é a alegação repetida, e o ranking responde a uma pergunta diferente da do contador: “quais falsidades ele mais repetiu?”, e não “quantas declarações falsas ele fez?”. Em 2023, a mais repetida foi a de que herdara 14 mil obras paralisadas (21 vezes; classificada como falsa) (Ribeiro, 2023). Em 2024, foi a de que o governo do PT acabou com a fome em 2014 (ao menos 24 ocasiões; classificada como falsa), e outra, repetida em ao menos 17 ocasiões, foi a de que seria o único presidente da República sem diploma uni-

versitário (classificada como falsa) (Menezes, 2024). Para Bolsonaro, o balanço de dez anos informa que a alegação mais repetida foi a de que não houve nenhuma denúncia de corrupção em seu governo (Menezes & Faustino, 2025).

Os dois conjuntos só se tornariam comparáveis se os dois presidentes fossem submetidos às duas perguntas: o ranking de alegações repetidas e o total de declarações checadas, no mesmo período e pelo mesmo critério.

■ DADO	O Aos Fatos publicou 6.675 declarações “falsas ou distorcidas” de Bolsonaro no mandato (dez./2022) e 6.685 em 2025; para Lula, publicou rankings anuais de alegações repetidas e classificadas como falsas. Nenhum dos dois traz o total de declarações checadas.
■ INTERPRETAÇÃO	São descrições fiéis da contagem de cada série, mas de unidades diferentes (declaração e alegação repetida) e sem denominador. Não permitem dizer quem “falou mais falsidades” nem quem as disse em maior proporção.
■ HIPÓTESE	Que o contador de Bolsonaro tenha crescido em ritmo parecido com a exposição (entrevistas longas, lives) é plausível pelo caso do Flow Podcast, mas não foi testado.

2.6 Falas e menções: de quem é a falsidade

Um erro de unidade frequente é tratar como equivalentes dois tipos de peça: a checagem do que o político disse e a verificação de um conteúdo que circula sobre ele. A Lupa os separa. “Checagem” é a verificação de falas, declarações ou afirmações específicas, geralmente atribuídas a pessoas públicas; “verificação”, a de conteúdos virais ou potencialmente virais, sem autoria identificada (Agência Lupa, 2026).

O Aos Fatos informa, em seu balanço de dez anos, que Lula e Bolsonaro foram as figuras públicas mais checadas e também as mais citadas em peças de desinformação: 677 menções para Lula e 566 para Bolsonaro. E acrescenta que, “enquanto o petista é mais atacado pelas peças enganosas, Bolsonaro, em maior parte, tem suas ações infladas” (Menezes & Faustino, 2025). Kniess et al. (2024) constataram que mais da metade das 119 checagens que compunham seu corpus destinava-se a refutar boatos desfavoráveis a Lula.

Esses números descrevem o que circula *sobre* cada político, não o que cada um disse. Contar peças de desinformação que citam um político como se fossem falas dele é um erro de unidade. Nos capítulos seguintes, trabalha-se apenas com falas atribuídas ao próprio político, e o conteúdo sobre ele é tratado como outro objeto.

2.7 As agências concordam entre si?

Uma checagem é uma classificação, e classificações podem divergir. Há três fontes de variação.

A primeira é a escala: a da Lupa mudou duas vezes (seção 2.4). A segunda é o conteúdo: Recuero et al. (2025), em 811 checagens de três agências brasileiras nas eleições de 2018 e 2022, encontraram uma transição de conteúdos completamente falsos para enganosos e de conteúdos sem fontes para conteúdos que usam fontes reais. O que se classifica mudou, e com isso a distribuição das etiquetas. Há estudos empíricos sobre a consistência entre checadores em debates eleitorais (e.g., Lelo, 2023), mas cujos dados pontuais não são incorporados a este cômputo (ver Apêndice D).

Enquanto a terceira pergunta não puder ser respondida com dados, cada classificação é tratada como atribuída à agência que a fez, e não como fato estabelecido.

2.8 O que a checagem não alcança

A checagem opera sobre enunciados factuais, que se comparam com registros: dados oficiais, documentos, gravações. A Análise do Comportamento oferece uma distinção útil para saber onde isso termina.

CONCEITO EM USO

Crença e atitude como comportamento verbal (Guerin, 1994)

Guerin trata atitudes e crenças como comportamento verbal moldado por comunidades verbais, e não como relatos de um estado interno. Uma *crença* (“reatores nucleares são inseguros”) pode ser contestada com evidência ou pelo enunciado contrário. A *atitude* correspondente (“não gosto da energia nuclear”) não pode ser atacada do mesmo modo: o falante pode responder, na formulação de Guerin, que, “apesar de tudo o que você disse”, continua não gostando (tradução minha). Por isso, apresentar um tema como atitude, e não como crença, é frequentemente reforçado, e a diferença entre as duas formas é uma contingência da comunidade de ouvintes, não uma propriedade do enunciado (Guerin, 1994).

CONCEITO EM USO

Autoclítico (Guerin, 1994)

Autoclíticos são qualificações acrescentadas ao enunciado que modificam o efeito dele sobre o ouvinte, como “parece-me que...”, “sempre achei que...” ou “gosto um pouco de cachorros” dito a uma sociedade de amantes de gatos (Guerin, 1994; termo de Skinner, 1957, citado por Guerin). Podem reduzir consequências negativas e permitem requalificar o que foi dito depois de uma réplica negativa.

O que as agências avaliam restringe-se a enunciados factuais — equivalentes, no vocabulário analítico-comportamental de Guerin (1994), à classe das crenças, as quais mantêm correspondência com estímulos antecedentes verificáveis. Em contrapartida, declarações formuladas como atitude (a exemplo de 'este governo é o pior da história') ou mediadas por qualificadores autoclíticos ('a meu ver', 'estou apenas indagando') encontram-se predominantemente sob controle de consequências sociais imediatas da audiência, situando-se fora do escopo de aplicação das etiquetas de veracidade factual. Ademais, Guerin (1994) demonstra que enunciados sob a forma de atitude frequentemente operam como mandos: a supervisora que declara em público desgostar de funcionários que tomam café durante o expediente exerce funcionalmente controle sobre o comportamento do ouvinte para que ele cesse a ação, independentemente de qualquer intenção privada.

A consequência para a leitura das agências é direta. A etiqueta mede a correspondência entre o que se disse e os registros. Não mede a função do que se disse na audiência. Dois enunciados, um checável e outro uma atitude ou uma qualificação autoclítica, podem ter efeitos semelhantes sobre quem ouve, e só o primeiro entra na contagem. Os Capítulos 3 e 5 analisam a função; este capítulo se limita a registrar que a parte do discurso que migra da forma factual para a opinativa sai do alcance da checagem sem que a função mude.

■ DADO	Guerin (1994) descreve que crenças podem ser contestadas com evidência e atitudes não, e que a apresentação sob a forma de atitude é frequentemente reforçada. A Lupa e o Aos Fatos checam enunciados factuais.
■ INTERPRETAÇÃO	Um enunciado que passe da forma factual para a opinativa deixa de ser checável, ainda que o que ele faz com o ouvinte seja parecido.
■ HIPÓTESE	Que parte dessa migração seja uma resposta às próprias contingências da checagem (esquiva de contracontrole, Capítulo 6) é uma hipótese a ser testada com falas datadas, antes e depois de checagens; não há dado sobre isso neste capítulo.

2.9 Protocolo para uma amostra comparável

Para comparar os dois presidentes sem que a escolha dos casos dependa do resultado, a regra abaixo foi fixada antes de qualquer coleta.

1. **Agência.** Uma só, a Lupa, que tem escala única e publica denominador. Misturar agências misturaria escalas e critérios de seleção.
2. **Janelas.** Os primeiros 45 meses de cada mandato: de 1º de janeiro de 2019 a 30 de setembro de 2022 (Bolsonaro) e de 1º de janeiro de 2023 a 30 de setembro de 2026 (Lula), por data de publicação da checagem. As duas têm a mesma duração e terminam antes do primeiro turno (2 de outubro de 2022 e 4 de outubro de 2026).
3. **Unidade.** A frase atribuída ao próprio político em uma checagem. Ficam de fora os conteúdos virais sobre ele e as falas de terceiros. Quando uma peça traz várias frases, conta-se cada uma; quando a marcação não as separa, conta-se a peça e sinaliza-se.
4. **Escala.** Mantêm-se as etiquetas originais. “Verdadeiro, Mas...” (antes de 2023) e “Falta Contexto” (depois) são agrupadas, como faz a Lupa; “De olho” e “Ainda é cedo” ficam à parte, como “sem veredito”.
5. **Saídas.** Número de frases por etiqueta e proporção, com intervalo de Wilson, para cada presidente; tabela separada por peças de frase única e de várias frases.
6. **Sensibilidade.** Repetir excluindo os períodos de campanha oficial, excluindo debates e entrevistas longas, e por trimestre, para ver se a diferença aparece ou some.

7. **Validação.** Comparar as contagens das campanhas de 2018 e 2022 com os 102 de 236 do relatório, e as de Lula com os 22 de 83, para estimar a cobertura da coleta.
8. **Limites declarados desde já.** A seleção é da agência; não há denominador de falas totais; a exposição dos dois varia ao longo do mandato; a coleta depende da cobertura da fonte.

A coleta não foi feita, por três razões. (a) Os arquivos robots.txt da Lupa e do Aos Fatos restringem a coleta automatizada por ferramentas computacionais, parâmetros que foram rigorosamente respeitados. (b) O repositório público ClaimReview do Google Data Commons, que agrega checagens estruturadas com esse metadado, apresenta cobertura histórica irregular em relação à Lupa: registra 3.369 checagens concentradas entre 2022 e setembro de 2024 (com apenas 12 entradas em 2019, 3 em 2020 e 4 em 2021, inexistindo registros sob o domínio atual), inviabilizando a delimitação de janelas temporais simétricas. (c) O relatório da Lupa não discrimina os ciclos de Lula. Os caminhos abertos são: a Lupa ceder a base do Observatório (a base que sustenta o relatório existe), ou a coleta ser feita pelo autor, sob sua responsabilidade.

Síntese

Uma fração publicada por uma agência é a proporção entre os enunciados que passaram por duas séries de seleção, a do político e a da agência (Figura 2.1). Por isso, antes de comparar números, convém perguntar de cada um qual é a unidade, qual é o denominador, qual é o período e a situação, qual é a escala e quem selecionou o que entrou (Quadro 2.1).

Feitas essas perguntas, os dados publicados hoje não sustentam a comparação entre Bolsonaro e Lula. A Lupa publica numerador e denominador por político (Bolsonaro, 102 de 236, ou 43,2%; Lula, 22 de 83, ou 26,5%), mas são frações de campanhas, com ciclos diferentes e não discriminados, e a etiqueta “Falso” é uma entre várias. O Aos Fatos publica, para Bolsonaro, um numerador sem denominador (6.675 em 2022; 6.685 em 2025) e, para Lula, rankings de alegações repetidas: são unidades diferentes. E as peças sobre um político, as menções, não são falas dele.

A checagem, por fim, incide estritamente sobre enunciados factuais (crenças, na conceituação de Guerin, 1994). Formulações de atitude e modulações auto-clíticas permanecem imunes à verificação, ainda que possam exercer funções comportamentais análogas sobre a audiência. Uma comparação legítima entre Bolsonaro e Lula exigiria o escrutínio de uma mesma agência, balizado por janelas temporais simétricas de exercício de mandato e acompanhado do res-

pectivo denominador. Como tal base probatória inexistente nas publicações correntes, a regra metodológica para sua eventual construção encontra-se formulada na seção 2.9.

Referências

- Agência Lupa. (2026, 27 de julho). *Anatomia da desinformação eleitoral* [Relatório]. https://uploads.agencialupa.org/2026/07/RELATORIO-ANATOMIA-DA-DESINFORMACAO-ELEITORAL_LUPA-FINAL.pdf
- Guerin, B. (1994). Attitudes and beliefs as verbal behavior. *The Behavior Analyst*, 17(1), 155–163. <https://doi.org/10.1007/bf03392661>
- Kniess, A. B., Comel, N., & Marques, F. P. J. (2024). Checagem de fatos e autoridade jornalística no Brasil: uma análise do Fato ou Fake e do Estadão Verifica. *Comunicação e Sociedade*, 46, Article e024020. [https://doi.org/10.17231/comsoc.46\(2024\).5665](https://doi.org/10.17231/comsoc.46(2024).5665)
- Lelo, T. (2023). Assessing the consistency of fact-checking in political debates. *Journal of Communication*, 73(6), 587–600. <https://doi.org/10.1093/joc/jqad028>
- Menezes, L. F. (2024, 23 de dezembro). *No segundo ano de governo, Lula desinformou mais sobre fome e educação*. Aos Fatos. <https://www.aosfatos.org/noticias/lula-2024-desinformou-fome-educacao/>
- Menezes, L. F., & Faustino, M. (2025, 7 de julho). *Como a desinformação transformou o debate político brasileiro nos últimos dez anos*. Aos Fatos. <https://www.aosfatos.org/noticias/desinformacao-transformou-debate-politico-dez-anos/>
- Recuero, R., Tavares, C., Volcan, T., Pozzebon, M., & Dutra, M. (2025). A desinformação narrada pela checagem: estudo das eleições de 2018 e 2022. *E-Compós*, 28, 1–26. <https://doi.org/10.30962/ecomps.3039>
- Ribeiro, A. (2023, 27 de dezembro). *Em primeiro ano de mandato, Lula repete alegações enganosas sobre fome, economia e governo Bolsonaro*. Aos Fatos. <https://www.aosfatos.org/noticias/lula-alegacoes-enganosas-2023/>
- Ribeiro, A., & Menezes, L. F. (2022, 22 de dezembro). *Uma retrospectiva das mentiras do governo Bolsonaro* [Matéria em quadrinhos]. Aos Fatos. <https://www.aosfatos.org/noticias/retrospectiva-mentiras-governo-bolsonaro/>
- Santos, L. F. dos. (2022, 17 de agosto). *Bolsonaro supera 6.000 declarações falsas ou distorcidas no início da campanha pela reeleição*. Aos Fatos. <https://www.aosfatos.org/noticias/bolsonaro-supera-6000-declaracoes-falsas-ou-distorcidas-no-inicio-da-campanha-pela-reeleicao/>
- Skinner, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts.



PARTE II

Análise operante dos discursos

-
- 3 Declarações falsas como operantes verbais
 - 4 Atitudes políticas e preconceito como classes de equivalência
 - 5 Discurso de ódio: mando disfarçado e controle aversivo
 - 6 Bolhas digitais como comunidades verbais e a esquiva de contracontrole
 - 7 O voto como comportamento verbal

3

Declarações falsas como operantes verbais

Em 28 de agosto de 2022, num debate entre candidatos à Presidência transmitido por Band, Folha, TV Cultura e UOL, Luiz Inácio Lula da Silva disse: “O menor desmatamento da Amazônia foi feito no meu governo”. A Agência Lupa classificou a frase como “Falso”: a menor taxa de desmatamento da Amazônia Legal registrada no sistema Prodes, do Inpe, foi a de 2012, no governo de Dilma Rousseff, com 4,6 mil km², a menor da série iniciada em 1989 (Agência Lupa, 2026). No mesmo debate, Lula disse ter orgulho de ter indicado “o primeiro negro para a Suprema Corte”; a Lupa também a classificou como “Falso”, porque antes de Joaquim Barbosa, indicado em 2003, dois outros ministros negros haviam integrado o Supremo Tribunal Federal, Pedro Lessa e Hermenegildo Rodrigues de Barros (Agência Lupa, 2026). Em 29 de setembro de 2022, num debate da TV Globo, Jair Bolsonaro disse: “Nenhum país do mundo comprou vacina em 2020”; “Falso”, segundo a Lupa, que cita um levantamento do Poder360 segundo o qual ao menos 50 países começaram a vacinar em 2020. Já em 3 de agosto de 2018, em entrevista ao Central das Eleições, da GloboNews, ele disse: “Eu nunca fui homofóbico”; “Falso”, na checagem da Lupa, que lembra a condenação por dano moral coletivo confirmada pelo Tribunal de Justiça do Rio de Janeiro em novembro de 2017, por declarações que a agência descreve como de teor homofóbico.

Nas quatro frases há um traço que não é a falsidade. Todas têm a forma gramatical de quem constata um fato, e todas trazem um dos termos que a própria Lupa lista entre os que tornam uma afirmação absoluta: “menor”, “primeiro”, “nenhum”, “nunca” (p. 41). A pergunta que costuma vir em seguida é se quem as disse mentia, se acreditava nelas ou se errou. Este capítulo troca a pergunta. Em vez de inferir o que o falante tinha em mente, pergunta que variáveis controlam a emissão de uma frase assim e o que a forma gramatical dela esconde. É o capítulo-núcleo do livro: apresenta a ficha de análise que os capítu-

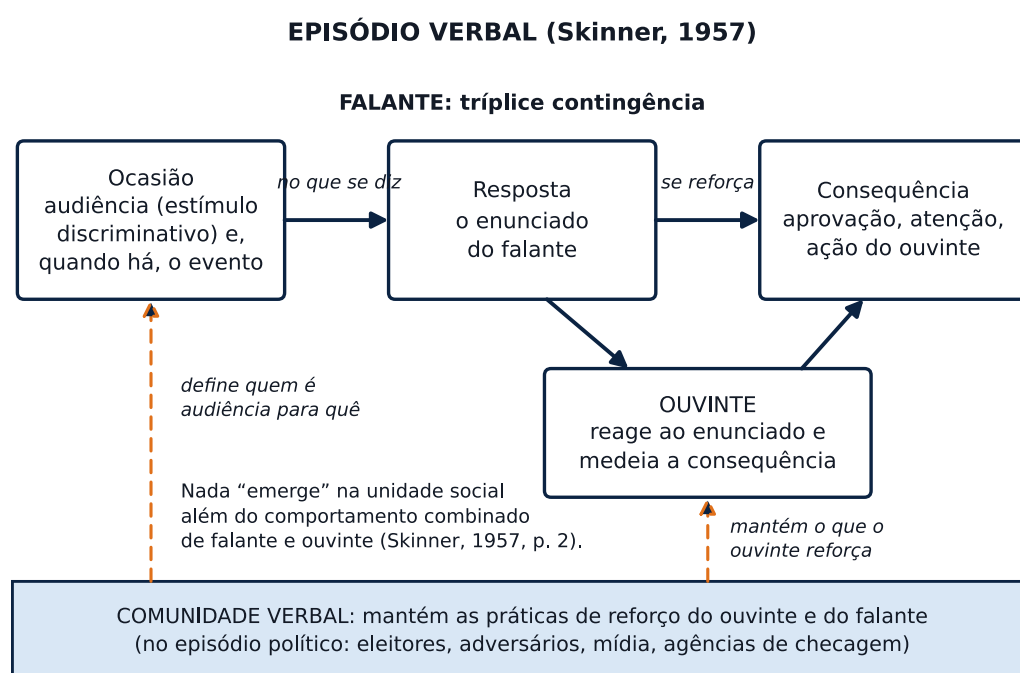
los seguintes usam e a aplica a esses quatro casos. Eles não foram escolhidos ao acaso nem são amostra: são os exemplos que a própria Lupa reproduziu no relatório para ilustrar o estilo de cada político (Agência Lupa, 2026), e entram aqui porque estão documentados em fonte lida, com data, contexto e a classificação atribuída a quem a fez.

3.1 A fala do político como comportamento verbal

Skinner (1957) define comportamento verbal como um tipo de comportamento necessariamente reforçado pela mediação de outras pessoas. A definição tem consequências práticas. O que um político diz não muda diretamente o mundo: muda o que os ouvintes fazem, e é por meio do que eles fazem que o comportamento do falante é conseqüenciado. Skinner chama de episódio verbal total o conjunto dos comportamentos de falante e ouvinte, e acrescenta que nada “emerge” na unidade social além do comportamento combinado de dois ou mais indivíduos” (p. 2). A Figura 3.1 reúne o esquema.

FIGURA 3.1

O episódio verbal: a tríplice contingência do falante, mediada pelo ouvinte e mantida pela comunidade verbal



Nota. Esquema original do autor, a partir de Skinner (1957). Não representa dados. Na ocasião, a audiência entra como estímulo discriminativo.

Três consequências dessa definição orientam o resto do capítulo. A primeira é que a forma gramatical não identifica o operante. Skinner nota que termos como “declarar” ou “afirmar” aplicados ao tato se referem a processos adicionais, os autoclíticos, e que o tato, tomado em si, não é por isso uma asserção (Skinner, 1957). Uma frase no indicativo, com sujeito e predicado, é apenas uma frase no indicativo, e só a história e as variáveis presentes dizem se foi controlada por um evento, por outras palavras ou pelo que o ouvinte faria depois. A segunda é que a fala de um político nunca tem uma só audiência. Skinner trata de audiências múltiplas e afirma que uma mesma resposta verbal pode produzir efeitos distintos a depender da audiência na presença da qual é emitida. Quando cada audiência é apresentada separadamente, o controle sobre o comportamento verbal é muito mais preciso. Por outro lado, quando há audiências distintas concomitantes, é comum a produção simultânea de consequências conflitantes. Num debate televisionado há os eleitores que o candidato quer convencer, os que já o apoiam, os adversários no palco e, depois, quem checa. Esses diferentes ouvintes se constituem em audiências conflitantes.

CONCEITO EM USO

Tríplice contingência, função × topografia e episódio verbal (Skinner, 1957)

Um operante verbal é definido pelo que controla a resposta, e não pela forma dela. Para Skinner (1957), nenhuma resposta pode ser classificada como um mando só pela sua forma e que, em geral, para identificar qualquer operante verbal é preciso conhecer o tipo de variáveis de que a resposta é função. A mesma topografia de resposta, isto é, a forma (e.g., “a vacina é perigosa”) pode, portanto, ser um tato, um intraverbal ou um mando, conforme a ocasião, a história e a consequência. Em outras palavras, as variáveis de controle da resposta verbal “a vacina é perigosa”. A *tríplice contingência* é a unidade: ocasião (estímulo discriminativo), resposta, consequência. No comportamento verbal, a ocasião inclui a audiência, que é um estímulo discriminativo em cuja presença o comportamento verbal é caracteristicamente reforçado, e a consequência vem do ouvinte. O *epísódio verbal* inclui as interações entre falantes e ouvintes cujas respostas verbais exercem múltiplas funções de estímulo sobre os comportamentos um do outro.

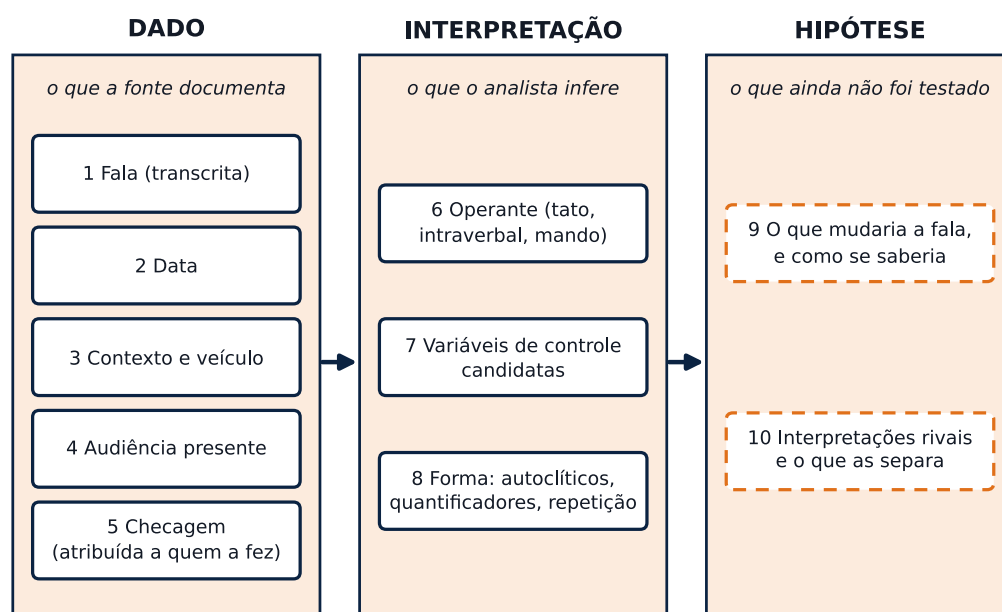
3.2 A ficha de análise

Toda análise deste livro passa pela mesma ficha e ela existe para obrigar a separar o que se sabe do que se infere. A Figura 3.2 mostra os campos. Os cinco primeiros são dados: a fala, transcrita como a fonte a traz; a data; o contexto e o veículo; a audiência presente, quando a fonte a documenta; e a checagem,

sempre com o nome de quem classificou, porque “falso” não é propriedade da frase, é decisão de uma agência com um método (Capítulo 2). Os três seguintes são interpretação: o operante provável, as possíveis variáveis de controle candidatas e os elementos verbais adicionais (autoclíticos, quantificadores, repetição). Os dois últimos são hipótese: o que mudaria a fala e como se saberia, e as interpretações rivais com o que as separa. Onde a fonte não documenta, a ficha escreve “não documentado” e não preenche o campo por inferência.

FIGURA 3.2

A ficha de análise: o que é dado, o que é interpretação e o que é hipótese



Em cada campo vazio, a ficha escreve “não documentado”; não o preenche por inferência.

Contorno tracejado: o que, nos casos do capítulo, é hipótese e não resultado.

Nota. Esquema original do autor. A versão para uso, com os campos em branco, está no Apêndice A.

Uma regra de leitura acompanha a ficha e é do próprio Skinner. Diante de registros do comportamento de um falante em que as condições de emissão são desconhecidas, é possível reconstruir uma história verbal provável, mas só raramente as inferências sobre as variáveis relevantes podem ser checadas diretamente. O critério de que se entende uma análise causal do comportamento verbal é a capacidade de prever a ocorrência de instâncias e, eventualmente, de produzi-las ou controlá-las. Os casos deste capítulo são registros de falas, emitidas por terceiros, sem acesso às condições em que foram emitidas;

por esse critério, suas análises são meras interpretações. É o que vale chamar de interpretação: uma leitura plausível, que diz quais variáveis poderiam controlar a resposta verbal e o que se precisaria observar para decidir entre elas.

A Tabela 3.1 reúne os operantes que a ficha usa, com a variável de controle de cada um (Skinner, 1957) e um exemplo construído para este capítulo, que não vem de nenhuma fonte e serve só para fixar a diferença.

TABELA 3.1

Operantes verbais: o que controla a resposta e um exemplo construído em fala política

Operante	O que controla a resposta (Skinner, 1957)	Exemplo construído
Mando	“Privação ou estimulação aversiva do falante; a resposta é reforçada por uma consequência característica e não tem relação especificada com estímulo anterior” (p. 35-36)	“Votem em mim.”; “Chega de enrolação.”; “Vamos mudar o Brasil”
Tato	“Objeto, evento ou propriedade de objeto ou evento: na presença dele, uma resposta de dada forma é caracteristicamente reforçada na comunidade verbal” (p. 81-82)	“O candidato empurrou o adversário durante o debate”, dito diante do ocorrido
Ecóico	“Estímulo verbal auditivo; a resposta gera padrão sonoro semelhante ao do estímulo” (p. 55). Há correspondência fonética entre estímulo e respostas vocais	Repetir (“Bandido bom é bandido morto”), ao responder, a expressão exata usada pelo adversário (“Bandido bom é bandido morto”)
Textual	“Estímulo verbal não auditivo (texto); resposta vocal sob o controle dele” (p. 66). Emissão de respostas vocais com fonemas correspondentes a estímulos escritos	Ler em plenário o trecho do projeto
Transcrição	“Estímulo verbal, com correspondência ponto a ponto entre sistemas dimensionais diferentes” (p. 71). Cópia: emitir uma resposta escrita com letras correspondentes ao estímulo escrito. Ditado: emitir uma resposta escrita com letras correspondentes aos fonemas presentes no estímulo vocal	Cópia: Digitar no celular uma tag correspondente à gerada por uma IA no computador. Tomar ditado: Tomar nota de afirmações de um adversário.
Intraverbal	“Estímulo verbal sem correspondência ponto a ponto com a resposta” (p. 71-72)	Responder “jamais será vencido” diante do início da palavra de ordem “O povo unido”

Nota. As definições são de Skinner (1957), nas páginas indicadas; a coluna de exemplos é do autor. Algumas categorias foram complementadas pelo autor, uma vez que a definição mais precisa do Skinner se dá a partir de fragmentos em parágrafos. Os seis tipos são operantes definidos pelas variáveis de controle, não por formas: a mesma frase pode pertencer a mais de um tipo em ocasiões diferentes.

3.3 Tato ou intraverbal? O contínuo e a forma de tato

Apenas quanto ao tato faz sentido avaliar a veracidade ou não da “resposta verbal” (dito, escrito ou gesticulado), uma vez que a sua variável de controle é um estímulo não verbal, ou seja, eventos do mundo. Skinner o define como o operante “em que uma resposta de dada forma é evocada por um objeto ou evento porque, na presença desse objeto ou evento, uma resposta dessa forma é caracteristicamente reforçada (fortalecida) na comunidade verbal” (p. 81–82). A comunidade verbal o estabelece porque os tatos aumentam o contato dos ouvintes com o ambiente. Quando alguém afirma que um candidato a governador deu uma cadeirada no outro durante um debate, está emitindo um tato desde que de fato a cadeirada tenha ocorrido. Paralelamente, um ouvinte que não assistiu ao debate terá acesso ao ocorrido (mais especificamente, terá acesso ao relato do ocorrido) mediante a emissão do tato pelo falante. No mando, por sua vez, o ouvinte tem acesso ao que pode reforçar a resposta verbal do falante, uma vez que o mando é o único operante verbal cujas respostas especificam o reforçador e, em muitos casos, a ação do ouvinte.

CONCEITO EM USO

Reforço social generalizado (Skinner, 1957; Guerin, 1994)

Qualquer evento que costuma preceder muitos reforçadores diferentes pode funcionar como reforçador para o comportamento sob todas as condições de privação e de estimulação aversiva apropriadas; um exemplo comum é a aprovação, que pode ser pouco mais que um aceno ou um sorriso de quem costuma fornecer uma variedade de reforçadores (Skinner, 1957). Como esse reforçador não depende de um estado particular de privação, a forma da resposta passa a depender do estímulo presente: é assim que se estabelece o tato, e é por isso que o reforço generalizado é, para Skinner, a chave do discurso prático e científico. Para Guerin (1994), as atitudes são enunciados mantidos por esse reforçamento social generalizado de uma comunidade verbal. Em decorrência de preceder diversos outros operantes passíveis de reforçamento, os reforçadores generalizados não guardam uma relação de correspondência com topografias de respostas específicas, ou seja, reforçam uma infinita multiplicidade de respostas. A aprovação, ilustrada pela fala “Muito bem!”, reforçará diferentes respostas, como resolver uma equação matemática ou pronunciar uma palavra corretamente em alemão, por exemplo.

A pureza do tato é, porém, uma questão de grau. Skinner chama de “puro” ou “objetivo” o tato mantido com reforço totalmente generalizado, o que permite o controle exclusivo pelo estímulo antecedente não verbal, o que raramente acontece. Em termos ideais, o discurso científico se preocupa em criar condições (método e rigor científicos) para que os tatos científicos se aproximem desse ideal. Mesmo no discurso científico, distorções ocorrem, uma vez que comunidades científicas reforçam certos relatos de resultados de pesquisa mais que outros. Fora a ausência de consenso epistemológico. No âmbito político, por sua vez, as distorções são muito mais prováveis, uma vez que estamos nos referindo ao controle sobre o comportamento das outras pessoas, no caso, dos ouvintes.

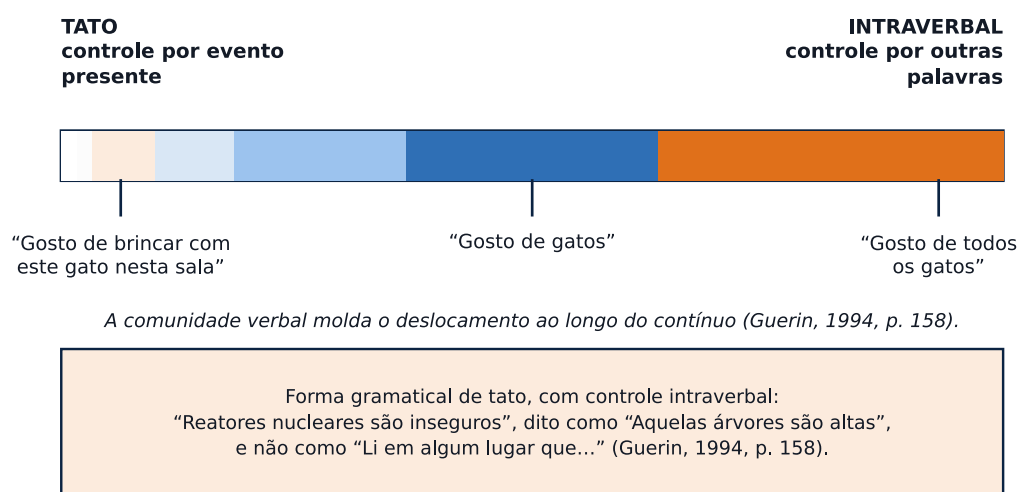
Já o intraverbal, por mais que seja reforçado pelo mesmo reforçador que tato, é controlado por estímulos verbais, os quais não guardam relação de correspondência ponto a ponto entre o estímulo e a resposta. A gigantesca maioria dos conteúdos ensinados na educação formal são intraverbais. Tomando a história, por exemplo, aprendemos sobre o saque do Rei João VI nos cofres brasileiros ao regressar a Portugal lendo ou ouvindo acerca desse evento histórico, uma vez que não o presenciamos. O que falamos ou escrevemos sobre esse evento são consideradas respostas intraverbais. Guerin (1994) leva essa ideia ao comportamento de relatar crenças e atitudes. Para ele, as distinções entre tatos e intraverbais é particularmente desafiadora em vários exemplos concretos, o que ainda não foi esclarecido a contento pela Análise do Comportamento. As comunidades verbais modelam o deslocamento ao longo de um contínuo, de “gosto de brincar com este gato nesta sala”, com alguma plausibilidade de tato, a “gosto de gatos” e “gosto de todos os gatos”, que são intraverbais. Muitas atitudes versam sobre eventos que a pessoa nunca experimentou, e, portanto, o que se sabe vem de jornais e de outras pessoas. A Figura 3.3 resume.

O que torna o intraverbal relevante para quem estuda desinformação é o que a comunidade reforça. Para Guerin (1994), as comunidades verbais reforçam a emissão de intraverbais como se fossem tatos (p. 155, 158): “reatores nucleares são inseguros”, na mesma forma gramatical de “aquelas árvores são altas”, é um enunciado mais poderoso e influente que “li em algum lugar que reatores nucleares são inseguros” (p. 158). Guerin (1992) analisa o conhecimento socialmente construído como intraverbais, tatos de discriminação frouxa e comportamento verbal sob fortes controles sociais do grupo. Uma mesma frase, sem autoclíticos que adicionem informações sobre as variáveis de controle (e.g., “andam falando por aí que...”), pode ser tanto um intraverbal como um tato a depender da variável de controle. Tatos, de modo geral, têm mais credibilidade dentre os ouvintes que os intraverbais. A sentença “Vi um determina-

do político casado entrando em um bordel” (i.e., autoclítico “vi” especifica a resposta verbal como um tato) tem mais credibilidade que “Me disseram que um determinado político casado entrando em um bordel” (i.e., o autoclítico “me disseram” especifica a respostas como um intraverbal). Por outro lado, a sentença “Um determinado político casado entrou em um bordel” pode ser tanto um tato como um intraverbal porque não contém elementos autoclíticos que informem ao ouvinte a sua variável de controle.

FIGURA 3.3

O contínuo tato–intraverbal e a forma de tato



Nota. Esquema do autor, com os exemplos de Guerin (1994). Guerin dá “gosto de gatos” e “gosto de todos os gatos” como intraverbais; a posição intermediária é ilustrativa. Não representa dados.

Um exemplo mostra o que isso significa para a análise. Mota et al. (2023) estudaram dez pronunciamentos de Bolsonaro sobre a vacina CoronaVac, entre julho de 2020 e janeiro de 2021. Escolheram, em buscadores, as dez falas de maior repercussão e de menção direta aos imunizantes e registraram de cada uma o número, a data, o contexto (veículo e ocasião) e o conteúdo transcrito. Classificaram as desordens informativas pela tipologia de Wardle e Derakhshan, em que desinformação é informação falsa criada deliberadamente para prejudicar e informação incorreta é informação falsa sem intenção de causar dano; e a má informação é a apresentação de informações verdadeiras, porém, fora de contexto para causar dano. O tipo mais frequente foi a informação incorreta (47,1%), seguida da desinformação (35,3%) e da má informação (17,6%). Numa das falas, em 29 de outubro de 2020, na “live” semanal, Bolsonaro disse que a vacina que menos tempo durou para ser colocada à disposição da população levou quatro anos. As autoras leem a frase como informação parcialmente verdadeira (de fato, a vacina de desenvolvimento mais rápido até en-

tão, contra a caxumba, havia levado quatro anos) usada em falso contexto, com a intenção, segundo elas, de pôr em dúvida as vacinas contra a covid-19, cujo desenvolvimento levou menos tempo. O que a ficha acrescenta à leitura das autoras é a separação entre o que o registro mostra e o que ela inferiu. O registro mostra uma frase que, tomada sozinha, corresponde a um dado verificável. Ele não mostra se a resposta foi controlada pelo fato (e então a falha estaria no contexto em que foi usada) ou por uma sequência de palavras aprendida como argumento (e então o fato seria tangencial). Dizer “falso contexto”, como Mota et al., é descrever o efeito da frase sobre o ouvinte, e não a variável que a controlou.

3.4 Tato distorcido: quando a consequência do ouvinte desloca o controle pelo evento

Skinner trata da distorção do tato no capítulo sobre as condições especiais que afetam o controle de estímulos (1957). O ponto de partida é que o reforço generalizado é a chave do discurso prático e científico: leva o comportamento do falante ao controle do ambiente corrente e permite ao ouvinte reagir ao que o falante diz no lugar do contato direto. Quando a correspondência com a situação estimuladora é nitidamente mantida, a resposta é chamada de objetiva, válida, verdadeira ou correta. Mas o controle de estímulos nunca é perfeito, e pode ser distorcido por consequências especiais que um ouvinte particular, ou os ouvintes em geral, fornecem, caso em que a resposta é chamada de subjetiva, preconceituosa ou desejosa. Quando o reforçamento generalizado é contingente a topografias de respostas verbais específicas correspondentes ou não aos estímulos a que se referem, tatos distorcidos tornam-se mais frequentes (Ribeiro, 1985). Se uma pessoa diz que um candidato a governador é corrupto, mesmo que não tenha nenhuma evidência para tal, provavelmente, opositores de tal candidato reforçaram essa resposta verbal com aprovação e suporte.

A descrição de Skinner é gradual e o gradiente importa. No caso menor, o falante exagera o tamanho do peixe que pescou ou minimiza o perigo de um ataque; a resposta ainda é controlada, em parte, pelo evento. Num caso maior, o controle é inventado, e a resposta é emitida quando o evento não ocorreu: a criança que perdeu uma moeda, disse que a perdeu e foi atendida passa a dizê-lo quando não perdeu nenhuma, e o menino que gritou “Lobo!” é o estereótipo. O exemplo mais claro do mecanismo é o do trovador. Ele começa relatando façanhas reais; certas partes recebem aprovação especial porque interessam ou lisonjeiam os ouvintes; essas partes sobrevivem nas narrações seguintes, e sob o mesmo reforço diferencial ele passa a esticar o relato, exagerando o tamanho da batalha e o heroísmo dos participantes, e, por fim, rompe

de todo o controle pelo estímulo, “descrevendo” cenas que nunca viu. A distorção é assim produto de seleção por consequências: o ouvinte reforça umas versões, o falante as repete e, passo a passo, a correspondência deixa de ser o que decide.

CONCEITO EM USO

Tato distorcido e subcomunidade verbal (Skinner, 1957; Guerin, 1992)

Tato distorcido é o tato cujo controle pelo evento foi deslocado por consequências especiais fornecidas pelo ouvinte. Skinner (1957, p. 149) descreve graus: o falante “estica os fatos”; o controle é “inventado”, e a resposta é emitida na ausência das circunstâncias em que é caracteristicamente reforçada; no limite, a resposta é emitida sob circunstâncias que controlam uma resposta incompatível, e a isso se chama mentira. Para Guerin (1992), o controle do tato pode estar nas mãos de uma *subcomunidade verbal* menor: se um pequeno grupo reforça certos relatos e pune outros, as características do grupo passam a controlar o que se relata, ainda que as consequências sejam generalizadas, e Skinner chamou esses tatos de distorcidos.

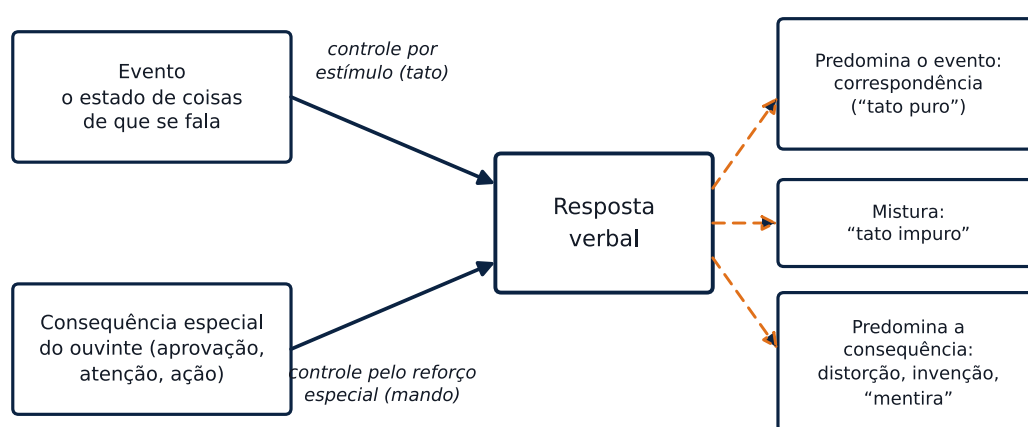
Skinner descreve ainda uma mistura que ocorre com frequência, o tato impuro. Quando uma dona de casa diz “O jantar está servido” principalmente porque os ouvintes então virão à mesa, a resposta é funcionalmente muito próxima do mando “Venham jantar!”; a mistura de relações de controle características do tato e do mando é o que ele chama de tato impuro. A Figura 3.4 mostra a lógica. O ponto é que uma resposta pode ser função das duas variáveis ao mesmo tempo, e as fontes de força se somam. Isso importa na interpretação das falas: mesmo quando o evento controla parte da resposta, a consequência do ouvinte pode controlar outra parte, e a falsidade, medida pela correspondência com o evento, não diz qual é a proporção.

Skinner acrescenta duas previsões que interessam à política. A primeira é a do custo para o sistema. Quando a distorção leva o ouvinte a reagir de modo ineficaz, o sistema falante–ouvinte pode se deteriorar: o ouvinte pode suspender o reforço ou punir as respostas do falante. O tato distorcido é “temporariamente mais eficaz” porque explora uma tendência maior do ouvinte a responder; a comunidade deixa de dar a moeda e pode punir a criança por mentir. A segunda é a da condição para a estabilidade: o sistema só é estável quando a correspondência com os estímulos de controle é de nenhuma importância prática para o ouvinte, como na literatura. E há a observação sobre a aversão: o controle de um tato é especialmente propenso à distorção quando a resposta é emitida para evitar ou escapar de consequências aversivas; a testemunha subornada é o exemplo.

Essas previsões foram escritas para uma díade (um falante, um ouvinte) e para uma comunidade em que o ouvinte reage à fala com ações que dependem de ela corresponder aos fatos. A fala política difere em três pontos, e é nesses pontos que a análise avança por hipótese. Primeiro, as audiências são várias, e quem reforça a distorção pode não ser quem a pune (seção 3.1); a deterioração prevista para a díade pode não ocorrer enquanto o comportamento do falante for reforçado por uma audiência e punido só por outra, sem acesso a ele. Segundo, para parte da audiência, a correspondência com os fatos pode ser, na terminologia de Skinner, de nenhuma importância prática, e a função da frase passa a ser outra: os efeitos que produzem nos ouvintes. Guerin (1994) nota que relatar atitudes pode servir de evento social ritual, em que o reforço está na própria troca. Terceiro, quando o falante pertence a uma subcomunidade que reforça certas respostas e pune outras, o que se tem é o que Guerin (1992) descreve: o que se relata passa a depender do grupo. Os Capítulos 6 e 9 tratam das duas últimas.

FIGURA 3.4

Controle múltiplo: o evento e a consequência especial do ouvinte atuam sobre a mesma resposta



Uma única resposta pode ser função das duas variáveis ao mesmo tempo; as fontes de força se somam (Skinner, 1957, p. 227-228).

Nota. Esquema do autor, a partir de Skinner (1957). Os três desfechos são tipos, não proporções; nenhuma fonte mede quanto da força de uma resposta política vem de cada variável.

■ DADO	Skinner (1957) descreve, sem medir, três propriedades do tato distorcido: ele é função de consequências especiais do ouvinte; é temporariamente mais eficaz que o tato; e leva o sistema falante-ouvinte a se deteriorar, salvo quando a correspondência não importa ao ouvinte. Guerin (1992) descreve o deslocamento do controle para pequenos grupos.
■ INTERPRETAÇÃO	A fala política pode ser lida pela mesma lógica, com as ressalvas de audiências múltiplas e de funções não informativas. Falas classificadas como falsas por uma agência são candidatas à análise de distorção, não casos demonstrados dela.
■ HIPÓTESE	Num campo em que a correspondência com os fatos tem pouca importância prática para a audiência que reforça, o tato distorcido não se deteriora e tende a se estabilizar. Nenhuma fonte verificada mede isso para falas de políticos.

3.5 Repetição: a força da afirmação repetida

A repetição aparece em duas posições e convém não confundi-las. Há a repetição do falante, que diz de novo a mesma frase em ocasiões diferentes (o que a configura como resposta intraverbal), e a repetição que o ouvinte sofre, quando encontra de novo a mesma frase. A primeira é variável dependente da análise operante: a frequência de emissão é medida de força da resposta. A segunda é variável independente em boa parte da psicologia da desinformação.

Do lado do falante, Skinner dá três pontos de apoio. O primeiro é a observação de que uma anedota ou piada particularmente bem-sucedida tende a ser contada de novo (1957, p. 148): medidas especiais de reforço “dizem ao falante sobre o que vale a pena falar” (p. 148). O segundo é o aviso de que a repetição pode ser nada mais que evidência de força excessiva da resposta, e não de uma variável de controle específica. O terceiro é que, quando o comportamento verbal é ineficaz em alterar as circunstâncias responsáveis por sua força, uma resposta pode ceder a outra, sobretudo quando a repetição pura é punida, e o que aparece é “repetição com variação” (p. 228). A predição é testável: se a repetição pura é punida, a forma deve variar. Isso ocorre muito em redação, na qual professores punem frases repetidas ao longo de um texto.

O que as fontes verificadas documentam para o Brasil é o número de vezes que certas alegações foram repetidas. O Aos Fatos registrou que a alegação de que os governos do PT teriam acabado com a fome no Brasil foi dita no discurso de posse de Lula, no parlatório do Palácio do Planalto, e repetida em ao menos outras sete ocasiões em 2023, e que era recorrente no discurso dele

desde a campanha (Ribeiro, 2023); em 2024, a alegação foi a mais repetida do ano, em ao menos 24 ocasiões, e foi classificada pelo Aos Fatos como falsa por comparar dois indicadores distintos (Menezes, 2024). A alegação de que seria o único presidente sem diploma universitário foi repetida em ao menos 17 ocasiões em 2024 e classificada como falsa (Menezes, 2024). No outro campo, a fonte equivalente é diferente em natureza. Mota et al. (2023) transcrevem duas falas de Bolsonaro, de 17 de dezembro de 2020 e de 15 de janeiro de 2021, que repetem, com palavras muito próximas, duas coisas: que ele já havia contraído o vírus e tinha anticorpos (“pra que tomar vacina de novo?”, “pra que tomar vacina?”) e que os laboratórios, nos contratos, não se responsabilizam por efeitos colaterais; e, segundo as autoras, que citam uma reportagem de O Globo de janeiro de 2021, ele fez ao menos dez críticas públicas à CoronaVac entre julho de 2020 e janeiro de 2021.

Quatro observações separam o que isso mostra do que não mostra. A primeira: as duas fontes medem coisas diferentes (o Aos Fatos conta ocasiões em que uma alegação foi repetida em um ano, a partir de discursos, entrevistas e transmissões ao vivo de 2023 e, em 2024, de todos os discursos, pronunciamentos e entrevistas do presidente; as autoras transcrevem dez falas escolhidas por sua repercussão), e por isso as contagens não se comparam. A segunda: a repetição da alegação sobre a fome continuou em 2024 depois de o levantamento de 2023 já a ter classificado como falsa (Ribeiro, 2023; Menezes, 2024); isso é um dado que não diz que Lula conhecesse a classificação, nem que a audiência a conhecesse. A terceira: o que se registra é a emissão, e não o que mantém a emissão. A quarta: as repetições registradas aqui são, na superfície, repetição pura, e o capítulo não sabe se em outras ocasiões a forma variou.

Para o lado do ouvinte, a fonte verificada é um conjunto de experimentos. Effron e Raj (2020) reuniram quatro experimentos e um piloto, dois deles pré-registrados, com 2.587 participantes no total, para perguntar se encontrar de novo uma manchete falsa muda o juízo moral sobre divulgá-la. No Experimento 1, 138 americanos de um painel online viram seis de 12 manchetes falsas quatro vezes, numa fase de familiarização em que avaliavam, a cada vez, um aspecto sem relação com a falsidade; depois julgaram as 12, seis já vistas e seis novas, informados de que verificadores não partidários haviam confirmado que aqueles fatos não aconteceram. Ver a manchete uma ou quatro vezes reduziu o quanto os participantes achavam antiético publicá-la e compartilhá-la quando a viam de novo, mesmo rotulada como falsa e desacreditada por eles, e mesmo depois de descontar estatisticamente o quanto ela parecia simpática e popular; e perceber a manchete como menos antiética previu maior inclinação a expressar aprovação a ela online. Num ambiente experimental, as pessoas compartilharam mais as manchetes repetidas que as novas (Effron & Raj,

2020). Os autores especulam que a repetição pode reduzir a condenação moral por fazer a informação parecer intuitivamente verdadeira, e discutem outros mecanismos. Em termos operantes, o resultado é compatível com uma leitura: a repetição muda o que a audiência faz com a frase (condena menos, compartilha mais), isto é, muda a consequência que o falante recebe pela repetição. É leitura do capítulo, não conclusão dos autores. E o desenho (manchetes, participantes de experimentos, ambiente simulado) está longe de um discurso de campanha.

■ DADO	Uma alegação foi repetida em ao menos 24 ocasiões em 2024 por Lula (Menezes, 2024), depois de ser classificada como falsa em 2023 (Ribeiro, 2023); duas coisas ditas por Bolsonaro em 17/12/2020 reaparecem em 15/01/2021 (Mota et al., 2023). Em experimentos, repetição reduziu a condenação moral e aumentou o compartilhamento (Efron & Raj, 2020).
■ INTERPRETAÇÃO	A repetição é medida de força da resposta, e força é efeito de história de reforço; a literatura experimental sugere que a audiência tolera mais o que já ouviu. Os registros não identificam quais consequências mantiveram as repetições.
■ HIPÓTESE	A repetição por políticos é mantida, em parte, pela reação da audiência a ela, e deve variar na forma quando a repetição pura passa a ser punida. Para testar, é preciso o registro das reações da audiência em cada ocasião e da forma da alegação antes e depois da checagem.

3.6 Autoclíticos de fonte e de certeza: o que se acrescenta e o que se retira

No Capítulo 2 o autoclítico apareceu como qualificação do enunciado que modifica o efeito dele sobre o ouvinte (Guerin, 1994). Skinner (1957, ca) o analisa por inteiro, e três famílias interessam aqui.

A primeira família é a dos autoclíticos de fonte. Skinner nota que perguntar “Você viu ou alguém lhe contou?” é perguntar se a resposta foi tato, ecoico ou intraverbal ao comportamento verbal de outra pessoa, e que, em muitas comunidades, é boa educação começar a falar com um autoclítico que indique de onde vem o que se diz; um autoclítico pode ainda indicar que o que se diz já é aceito pelos ouvintes ou pelas pessoas em geral. São o contrapeso do que a seção 3.3 descreveu: a comunidade reforça intraverbais com forma de tato, e o autoclítico de fonte é justamente o que diria que a resposta não é tato. A Figura 3.5 propõe a hipótese de que, ao longo de uma cadeia de retransmissão, os

autoclíticos de fonte e de fraqueza se perdem e o que sobra é a forma de tato; Skinner descreve que uma resposta atípica numa conversa tende a ser retomada e passada de falante a falante, mas nenhuma fonte verificada acompanha uma cadeia desse tipo.

CONCEITO EM USO

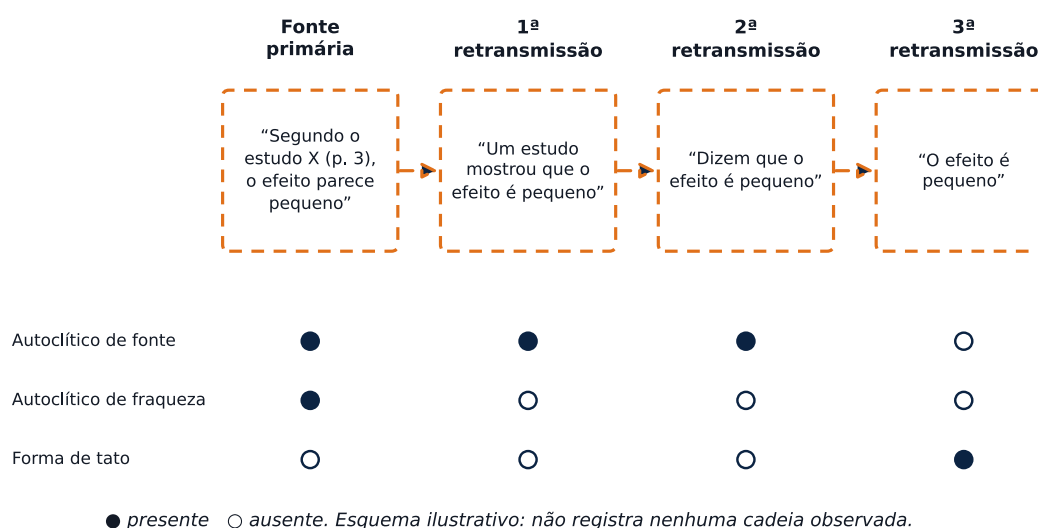
Autoclíticos de fonte e de certeza (Skinner, 1957)

O autoclítico é comportamento verbal baseado em outro comportamento verbal do mesmo falante e que modifica a reação do ouvinte a ele. Os *descritivos* informam o tipo de operante que acompanham (“vejo que...”, para o que se leu; “lembro”) e a força da resposta: “suponho” e “acredito” indicam estimulação insuficiente ou resposta mal condicionada (p. 315); “sei”, “garanto” e “juro” sugerem a força da resposta, e poderiam ser omitidos sem mudar a natureza do efeito, mas o tornam mais preciso. Há os que indicam a condição emocional do falante (“tenho o prazer de dizer”; p. 316) e os que apontam a fonte (“dizem que...”). Os *qualificadores* incluem o “não”, que tem força de mando e se traduz “não aja sobre esta resposta como um tato não estendido” (p. 324), e o autoclítico assertivo, que ordena ao ouvinte que aceite um estado de coisas. Os *quantificadores* (“todos”, “alguns”, “nenhum”) modificam a reação do ouvinte à sentença inteira.

FIGURA 3.5

Hipótese: perda dos autoclíticos de fonte e de fraqueza ao longo de uma cadeia de retransmissão

HIPÓTESE (sem dados): o que sobra do enunciado ao longo da retransmissão



Nota. Hipótese aqui proposta. As frases são construídas; o esquema não registra nenhuma cadeia observada e serve para dizer o que seria preciso observar: o enunciado em cada elo e os autoclíticos de cada um.

A segunda família é a dos autoclíticos de fraqueza e de certeza, e um exemplo está em uma das falas registradas por Mota et al. (2023). Em 24 de dezembro de 2020, Bolsonaro disse, em transmissão semanal: “A eficácia daquela vacina em São Paulo parece que está lá embaixo, né? Não vou divulgar percentual aqui, porque se eu errar 0,001% eu vou apanhar da mídia, mas parece que o percentual tá lá embaixo” (p. 320). As autoras classificam como falsa a informação de que o percentual estaria “lá embaixo” (p. 323). A frase traz duas vezes “parece que”, que é autoclítico de fraqueza no sentido de Skinner (indica resposta baseada em estimulação insuficiente), e uma referência explícita à audiência que pode puni-la: a mídia. A forma, portanto, informa ao ouvinte o grau de força da resposta, e a referência à mídia descreve a contingência que, segundo o próprio falante, o controlaria caso desse o número. A fala não tem a forma de tato que a seção 3.3 descreveu: ela se apresenta como palpite. Na hipótese da Figura 3.5, é um enunciado assim, com autoclítico de fraqueza, que perderia o autoclítico ao ser retransmitido; o registro de Mota et al. não mostra se isso ocorreu.

A terceira família é a mais pertinente para os dados da Lupa. Os termos absolutos (“sempre”, “nunca”, “todos”, “nenhum”, “único”, “primeiro”, “maior”, “menor”) e as negações fortes (“não existe”, “não há”) são, na classificação de Skinner, quantificadores e qualificadores: “todos” equivale a “sempre é possível dizer” e “nenhum” a “nunca é possível dizer” (p. 329), e o “não” tem força de mando. Eles definem a extensão com que se pede ao ouvinte que aceite a resposta, e Skinner nota, a propósito do autoclítico assertivo (e não dos quantificadores), que qualquer condição colateral que enfraqueceria a reação do ouvinte, como a negação por outra pessoa ou circunstâncias duvidosas, leva o falante a intensificá-lo.

A Lupa mediu a frequência desses padrões nas frases que checou em períodos eleitorais, e o resultado merece leitura cuidadosa. Segundo o relatório, afirmações com termos absolutos foram classificadas como “Falso” em 40,6% das checagens em período eleitoral, e frases com negações do tipo “não existe” ou “não há”, em 45,9% (Agência Lupa, 2026); as que mencionam o processo eleitoral, em 33,2%; as que atacam adversários, em 27,9%; e as que trazem números, percentuais ou valores, em 18,3%, embora estas recebam com mais frequência “Exagerado”, “Subestimado” ou “Verdadeiro, mas / falta contexto” (p. 50). Das 936 checagens classificadas como “Exagerado” ou “Verdadeiro, mas / falta contexto”, 737 (78,7%) continham algum marcador de escala, comparação, quantidade, desempenho, autoria ou legado. O texto não informa os numeradores e denominadores das porcentagens de 40,6%, 45,9%, 33,2% e 27,9%, nem se a base de cada porcentagem é o conjunto das checagens de cada padrão.

Há ao menos três leituras que não se excluem. A leitura lógica: uma frase universal ou negativa é refutada por um único contraexemplo, e uma frase com ressalva não é; a taxa de “Falso” seria, em parte, propriedade da forma lógica, e não do falante. A leitura de seleção: as agências checam o que é verificável e o que circula; a Lupa reconhece que o volume de checagens de cada político depende do tempo de exposição, das entrevistas e debates e, “sobretudo”, do estilo de comunicação, e que os dados refletem só o que a Lupa publicou na editoria de Eleições, condicionado a escolhas editoriais e de relevância. Já a leitura operante, que é aqui proposta: o autoclítico absoluto é emitido onde a reação do ouvinte precisa ser fortalecida, isto é, em ocasiões de disputa, de contestação ou de dúvida. A terceira só se testa com o registro da ocasião, que o relatório não traz.

3.7 A declaração como regra para a audiência

Uma declaração política pode ter outra função além de descrever um estado de coisas: descrever uma contingência, isto é, dizer ao ouvinte o que acontecerá se ele fizer isso ou aquilo. “Quem tomar essa vacina vai virar jacaré” descreve um evento (o efeito da vacina), mas, sobretudo, uma relação entre o comportamento do ouvinte e uma consequência. Na literatura analítico-comportamental, esse operante e o comportamento que ele evoca são tratados como regras e comportamento governado por regras é o comportamento do ouvinte em relação a elas.

CONCEITO EM USO

Regra e comportamento governado por regras (Baum, 1995; Cerutti, 1989)

Para Baum (1995), regra é um estímulo discriminativo verbal produzido pelo falante sob o controle de uma contingência, isto é, de uma relação entre o comportamento e suas consequências. Cerutti (1989) define o comportamento governado por regras como a combinação, em instruções complexas, de discriminações elementares estabelecidas antes. Afirma-se que o comportamento é governado por regras quando demonstra-se que a sua probabilidade de ocorrência é afetada pelo contato com uma regra. A propriedade que interessa aqui é a *insensibilidade às contingências*: desempenhos instruídos são menos sensíveis a mudanças nas consequências programadas (Hayes et al., 1986).

A literatura experimental disponível aqui, toda de laboratório, é consistente com a ideia de que a regra, em si, pode controlar o comportamento do ouvinte sem que ele contate a contingência. Em Galizio (1979), seis estudantes de 18 a 25 anos, recrutados por um anúncio de emprego de meio período em que o pagamento dependia do desempenho, giravam uma alavanca para evitar perdas

de dinheiro, sinalizadas por uma luz vermelha e um tom; quatro luzes âmbar, com rótulos, indicavam o componente do esquema múltiplo de esquiva em vigor. Instruções acuradas controlaram o desempenho em cada componente, e o seguimento de instruções foi eliminado quando instruções inexatas levaram os participantes a contatar uma contingência de perda monetária; enquanto a perda programada estava retirada, os participantes seguiram a instrução inexata, sem tendência de queda nas taxas. Hayes et al. (1986) apresentaram a participantes uma tarefa em que apertar botões movia uma luz por um labirinto, e mover a luz rendia pontos que valiam chances de prêmios em dinheiro; o movimento seguia um esquema múltiplo de razão fixa 18 e de reforçamento diferencial de taxa baixa de 6 segundos, com componentes que alternavam a cada 2 minutos. No segundo experimento, luzes pareadas com as instruções “Go Fast” e “Go Slow” (“vá rápido”, “vá devagar”) eram apresentadas e depois retiradas. Os autores apontam que desempenhos instruídos são menos sensíveis a mudanças nas consequências e propõem que parte dessa insensibilidade se deve a contingências adicionais introduzidas pelo controle instrucional, como as consequências sociais do seguir a regra.

Para a análise de falas, a consequência é de duas ordens. A primeira é que a falsidade de uma regra pode estar na descrição da contingência, e não na descrição do evento: “se você virar jacaré, é problema de você”, registrada por Mota et al. (2023) na fala de 17 de dezembro de 2020, tem uma parte descritiva (o efeito colateral) e uma parte de regra (a consequência para quem se vacina). As autoras a leem como provável recurso satírico ou de hipérbole, a observar junto ao potencial impacto da fala de um chefe de Estado. Pela ficha, o que se registra é a forma (regra, com consequência descrita) e o que se hipotetiza é a função: se o ouvinte que a segue contata a contingência real, o experimento de Galizio sugere que o seguimento se extingue; se não contata, ou se o seguir a regra tem consequências sociais, a insensibilidade descrita por Hayes e colegas torna o seguimento da regra resistente. A segunda é a dependência da audiência: a regra falsa resiste na medida em que o ouvinte não entre em contato com a contingência real e em que outras contingências, as sociais, mantenham o seguir a regra. Nenhuma fonte verificada mede isso com falas políticas.

3.8 Casos: as mesmas quatro frases, pelo mesmo método

A Tabela 3.2 aplica a ficha às quatro frases da abertura. Duas são de Lula e duas de Bolsonaro, todas dos exemplos que a Lupa reproduz no relatório, todas classificadas por ela como “Falso”, no mesmo método e na mesma publicação. A ficha repete, para cada uma, o que a fonte documenta e o que não documenta. A coluna de operante é interpretação, nos termos da seção 3.2.

TABELA 3.2

Quatro falas classificadas como “Falso” pela Lupa: dado, interpretação e o que a fonte não documenta

Caso	Fala, data, contexto e classificação (dado)	Interpretação (operante)	Não documentado
1. Lula	“O menor desmatamento da Amazônia foi feito no meu governo.” Debate Band, Folha, TV Cultura e UOL, 28/08/2022. Lupa: “Falso”; a menor taxa registrada (4,6 mil km ² , 2012) foi no governo Dilma Rousseff (PT)	Tato sobre o próprio governo, com superlativo; distorção na direção do crédito, ou intraverbal de uma narrativa sobre “os governos do PT”	Pergunta ou argumento a que respondia; reação do público; se o dado já fora checado
2. Lula	“Tenho orgulho de ter indicado o primeiro negro para a Suprema Corte.” Mesmo debate. Lupa: “Falso”; dois ministros negros haviam integrado o STF antes da indicação de Joaquim Barbosa, em 2003	Autoclítico emocional (“tenho orgulho”) mais um tato com ordinal; fato histórico, em geral adquirido como intraverbal, emitido com forma de tato	Idem; se havia pergunta sobre o tema; reação
3. Bolsonaro	“Nenhum país do mundo comprou vacina em 2020.” Debate TV Globo, 29/09/2022. Lupa: “Falso”; ao menos 50 países começaram a vacinar em 2020 (Poder360); no Brasil, a vacinação começou em 17/01/2021	Tato com quantificador universal negativo; função possível de tato impuro, conforme o que o ouvinte faria com a frase (hipótese; recorrente; reação a ocasião não está documentada)	A pergunta ou o argumento anterior; se a frase era
4. Bolsonaro	“Eu nunca fui homofóbico.” Entrevista ao Central das Eleições (GloboNews), 03/08/2018. Lupa: “Falso”; condenação por dano moral coletivo confirmada pelo TJRJ em nov./2017, por declarações feitas em 2011 no programa CQC	Tato do próprio comportamento verbal, com “nunca”; possível distorção sob contingência aversiva (hipótese)	A pergunta que a ocasionou; se a frase foi repetida; reação; resposta a correções

Nota. Agência Lupa (2026). As frases, a classificação e as justificativas estão em imagens das checagens reproduzidas no relatório (recortes em verificacao/evidencias/); a coluna de interpretação é do autor. “Falso” é a etiqueta que a Lupa aplica a informação comprovadamente incorreta.

Os casos de Lula têm em comum a primeira pessoa e o crédito: “no meu governo”, “tenho orgulho de ter indicado”. O relatório os apresenta como instâncias de um estilo, e diz que uma parcela expressiva das declarações “Falsas” de

Lula em campanha está relacionada à autoria ou ao pioneirismo de políticas públicas, programas, leis e marcos institucionais (Agência Lupa, 2026). Para a ficha, importa que, no caso 1, a justificativa da Lupa aponta o dado verdadeiro (a menor taxa da série) num governo do mesmo partido, de modo que o enunciado diverge do dado numa só variável: a quem se atribui. Há ao menos duas leituras. Numa, a consequência do ouvinte (aprovação por créditos) deslocou o controle: é o tato distorcido na direção do crédito. Noutra, o falante emitiu um intraverbal, aprendido dentro de uma narrativa sobre “os governos do PT”, e a primeira pessoa é só a forma que ela tomou naquele momento. Elas se distinguem por duas observações que o relatório não traz: o que o mesmo falante diz do mesmo dado perante audiências que não reforçam o crédito, e se a forma mudou depois da checagem. No caso 2, a frase tem duas partes: a justificativa da Lupa reconhece a indicação de Joaquim Barbosa por Lula em 2003 e aponta dois antecessores negros no STF, de modo que a falha está em “primeiro”. “Tenho orgulho de” é, no vocabulário de Skinner, um autoclítico que indica a condição emocional do falante e enfatiza a relação pessoal com ele; a ordinal é o ponto de falha.

Os casos de Bolsonaro, no relatório, ilustram outro estilo: a Lupa diz que ele costuma receber “Falso” quando usa números, percentuais e estatísticas, e quando faz afirmações categóricas e acusações contra adversários, com “nunca”, “jamais”, “nenhum” e “sempre” aparecendo com frequência. Ambos os casos têm quantificador absoluto. No caso 3, o contraexemplo é público e verificável (o Reino Unido começou a vacinar em 8 de dezembro de 2020, segundo a Lupa), o que é exatamente o que torna a forma universal arriscada: uma única exceção a refuta. No caso 4, a frase é um tato do próprio comportamento verbal, e Skinner (1957) escreve que as contingências necessárias ao comportamento autodescritivo são arranjadas pela comunidade quando ela tem razão para perguntar “O que você disse?” e “Por que disse isso?” (p. 314). Guerin (1994) acrescenta que uma pergunta pode mandar a resposta. A fonte não traz a pergunta, e é a pergunta que decide se a resposta é emitida sob a contingência de uma acusação, que é a condição em que, segundo Skinner, o tato é especialmente propenso à distorção. A ficha deixa o campo em branco, e a leitura como distorção sob contingência aversiva fica como hipótese.

Há outras falas de Bolsonaro documentadas com data, contexto e transcrição pelas dez que Mota et al. (2023) analisaram, e elas entraram nas seções 3.3, 3.5, 3.6 e 3.7. Entre as fontes verificadas, não há equivalente para Lula: nenhuma transcreve falas dele com data e contexto e as classifica com o mesmo esquema. Isso é limite do acervo de fontes lidas e não propriedade dos políticos, e deve ser lido assim: uma conclusão comparativa a partir dessa assimetria seria artefato da documentação. Mota et al. classificam as falas pela tipologia de

Wardle e Derakhshan (seção 3.3), em que a desinformação se distingue da informação incorreta pela intenção de causar dano. A classificação, nesse esquema, depende de uma inferência sobre intenção que o registro de uma fala não contém; a ficha a substitui por uma pergunta sobre variáveis de controle.

A Tabela 3.3 reúne o que o relatório da Lupa publica sobre os padrões de linguagem nas frases checadas de cada um, com a mesma taxonomia e o mesmo formato.

TABELA 3.3

Padrões de linguagem nas frases checadas pela Lupa em campanhas: Lula e Bolsonaro

Padrão na frase checada	Lula: frases (falsas; %)	Bolsonaro: frases (falsas; %)
Negações fortes	4 (3; 75,0%)	25 (15; 60,0%)
Termos absolutos	7 (4; 57,1%)	39 (21; 53,8%)
Ataques a adversários ou partidos, crime ou corrupção	12 (7; 58,3%)	52 (21; 40,4%)
Processo eleitoral, voto e urna	3 (1; 33,3%)	13 (6; 46,2%)
Números, percentuais ou valores	57 (7; 12,3%)	94 (37; 39,4%)
Total de frases checadas	83 (22; 26,5%)	236 (102; 43,2%)

Nota. Agência Lupa (2026, p. 25), “Taxonomia Lupa”, valores lidos na imagem da página (recorte em verificacao/evidencias/lupa2026-p25-estilo-lula-bolsonaro.png). As categorias não são excluídas: uma mesma frase pode conter número, termo absoluto e ataque ao adversário. Só entram políticos com mais de 30 frases checadas; as frases são de campanhas eleitorais, e a Lupa não discrimina os ciclos de Lula.

O capítulo registra o que a tabela permite e o que não permite. Permite dizer que, nas frases que a Lupa checkou, as que tinham negação forte, termo absoluto ou ataque a adversário receberam “Falso” em proporção maior que as que traziam números, nos dois políticos, e que, entre as frases com números, a proporção de “Falso” foi de 12,3% (7 de 57) para Lula e de 39,4% (37 de 94) para Bolsonaro. Não permite dizer que um deles tende mais que o outro a afirmar o que é falso: as bases são seleções da agência, com tamanhos muito diferentes, as categorias se sobrepõem, as células de Lula têm 3, 4, 7 e 12 frases, e a própria Lupa diz que não faz ranking de quem “mentiu mais”, porque o número de checagens depende do tempo de exposição, do cargo e do estilo de comunicação. Nesses termos, o dado é a classificação da agência; a interpretação é que a forma da frase correlaciona com a classificação; a hipótese é a da seção 3.6.

3.9 Interpretações rivais e como decidir entre elas

Para uma fala classificada como falsa, a ficha trabalha com seis interpretações rivais, que a Tabela 3.4 reúne. Não se inclui entre elas “intenção de enganar”, porque intenção não é variável observável de controle; o que há são as consequências ao comportamento do falante, a história de reforço dele e as propriedades da ocasião. A definição de mentira de Skinner é funcional: uma resposta emitida sob circunstâncias que normalmente controlam uma resposta incompatível (1957).

TABELA 3.4

Interpretações rivais para uma fala classificada como falsa e o que a ficha precisaria registrar para separá-las

Interpretação	Variável de controle em jogo	O que a ficha precisaria registrar
Controle fraco pelo evento (erro, lapso)	Estimulação insuficiente ou resposta mal condicionada; sem consequência especial	Se a mesma pessoa descreve o evento corretamente em outras ocasiões, sem consequência especial; se há autoclítico de fraqueza
Intraverbal adquirido de outros	Estímulo verbal anterior (texto, discurso, mensagem)	A fonte da cadeia de palavras; se o enunciado reproduz formulação anterior; ausência de autoclítico de fonte
Tato distorcido	Consequência especial do ouvinte	Se a forma muda com a audiência; se a correspondência é maior quando a audiência não reforça
Mando disfarçado ou tato impuro	Privação ou aversão do falante; o que o ouvinte fará depois	O que o falante pede ao ouvinte, de modo explícito ou implícito, na mesma ocasião (Capítulo 5)
Função ritual ou de identidade	Reforço social generalizado: a troca é o reforço	Se o conteúdo varia sem mudar o efeito; se a audiência reage à adesão, e não ao conteúdo (Guerin, 1994)
Efeito da seleção pela agência	A fala entrou na base por ser checável e circular	A base e o método de seleção da agência (Capítulo 2)

Nota. Quadro do autor, a partir de Skinner (1957) e Guerin (1994). As seis interpretações não se excluem: Skinner (1957) nota que uma resposta costuma ser função de mais de uma variável.

Há quatro tipos de observação que a ficha pediria e que os registros usados neste capítulo não trazem. O primeiro é a ocasião: a pergunta, o argumento ou o contexto imediato que a fala responde. O segundo é a audiência: quem estava presente e como reagiu, que é a consequência de que depende o tato distorcido. O terceiro é a comparação do mesmo falante e do mesmo evento perante audiências diferentes, ou antes e depois da checagem; é o único modo de sepa-

rar distorção, por consequência especial, de controle fraco pelo evento. O quarto é o que acontece depois da correção: se a forma permanece (repetição pura), varia (repetição com variação) ou desaparece. Os Capítulos 6 e 9 tratam do que as fontes dizem sobre a reação da audiência e sobre o que acontece depois da correção; os Capítulos 5 e 7 usam daqui a distinção entre tato, mando e intraverbal para analisar discurso de ódio e voto.

■ DADO	Quatro falas, duas de cada político, foram classificadas como “Falso” pela Lupa; uma série de falas de Bolsonaro sobre a vacina foi transcrita e classificada por Mota et al. (2023); duas alegações de Lula foram contadas como repetidas pelo Aos Fatos.
■ INTERPRETAÇÃO	As falas são compatíveis com mais de um operante, e a forma gramatical de tato e os quantificadores absolutos aparecem em todas as quatro. A classificação de falsidade informa a correspondência com um dado, e não a variável de controle.
■ HIPÓTESE	Parte das falas é função de consequências especiais da audiência (tato distorcido), parte de intraverbais aprendidos de outros e parte de funções não informativas (mando, ritual). Para separá-las é preciso registrar ocasião, audiência, comparação e depois da correção.

Síntese

Uma fala política é comportamento verbal e o que a identifica como tato, mando ou intraverbal é o que a controla, e não a forma. O tato corresponde a um evento, e o intraverbal a outras palavras; a comunidade reforça apresentar o segundo na forma do primeiro. O tato distorcido é o tato cujo controle pelo evento foi deslocado pela consequência especial do ouvinte, em graus que vão do exagero à invenção. A repetição é medida de força, e a literatura experimental sugere que a audiência condena menos o que já ouviu anteriormente. Os autoclíticos de fonte e de certeza dizem ao ouvinte de onde vem o que se diz e com quanta força; os absolutos e as negações pedem que ele aceite um estado de coisas, e a Lupa mostra que frases assim recebem “Falso” com mais frequência, por razões lógicas, de seleção e, talvez, operantes.

A ficha separa dado, interpretação e hipótese. Aplicada a quatro falas de dois políticos, mostra o que cada fonte documenta e o que não documenta, e deixa em aberto a pergunta que importa: o que mantém a fala. Os Capítulos 5 e 7 usam deste capítulo a distinção entre tato, mando e intraverbal; o Capítulo 6, o tato distorcido e a subcomunidade verbal; o Capítulo 9, o reforço social gene-

realizado e a audiência; o Capítulo 12 (seções 12.1 e 12.4), a ideia de que corrigir o enunciado é agir sobre uma resposta cujas variáveis de controle podem não ser as que o corretor supõe.

Referências

- Agência Lupa. (2026, 27 de julho). *Anatomia da desinformação eleitoral* [Relatório]. https://uploads.agencialupa.org/2026/07/RELATORIO-ANATOMIA-DA-DESINFORMACAO-ELEITORAL_LUPA-FINAL.pdf
- Baum, W. M. (1995). Rules, culture, and fitness. *The Behavior Analyst*, 18(1), 1–21. <https://doi.org/10.1007/bf03392688>
- Catania, A. C., Matthews, B. A., & Shimoff, E. (1982). Instructed versus shaped human verbal behavior: Interactions with nonverbal responding. *Journal of the Experimental Analysis of Behavior*, 38(3), 233–248. <https://doi.org/10.1901/jeab.1982.38-233>
- Cerutti, D. T. (1989). Discrimination theory of rule-governed behavior. *Journal of the Experimental Analysis of Behavior*, 51(2), 259–276. <https://doi.org/10.1901/jeab.1989.51-259>
- Effron, D. A., & Raj, M. (2020). Misinformation and morality: Encountering fake-news headlines makes them seem less unethical to publish and share. *Psychological Science*, 31(1), 75–87. <https://doi.org/10.1177/0956797619887896>
- Galizio, M. (1979). Contingency-shaped and rule-governed behavior: Instructional control of human loss avoidance. *Journal of the Experimental Analysis of Behavior*, 31(1), 53–70. <https://doi.org/10.1901/jeab.1979.31-53>
- Guerin, B. (1992). Behavior analysis and the social construction of knowledge. *American Psychologist*, 47(11), 1423–1432. <https://doi.org/10.1037/0003-066x.47.11.1423>
- Guerin, B. (1994). Attitudes and beliefs as verbal behavior. *The Behavior Analyst*, 17(1), 155–163. <https://doi.org/10.1007/bf03392661>
- Hayes, S. C., Brownstein, A. J., Zettle, R. D., Rosenfarb, I., & Korn, Z. (1986). Rule-governed behavior and sensitivity to changing consequences of responding. *Journal of the Experimental Analysis of Behavior*, 45(3), 237–256. <https://doi.org/10.1901/jeab.1986.45-237>
- Menezes, L. F. (2024, 23 de dezembro). *No segundo ano de governo, Lula desinformou mais sobre fome e educação*. Aos Fatos. <https://www.aosfatos.org/noticias/lula-2024-desinformou-fome-educacao/>
- Mota, A. A. S., Pimentel, S. M., & Oliveira, A. V. M. G. (2023). Desordens informativas: análise de pronunciamentos de Jair Bolsonaro contra a vacinação de covid-19. *Revista Eletrônica de Comunicação, Informação & Inovação em Saúde*, 17(2), 311–331. <https://doi.org/10.29397/reciis.v17i2.3513>
- Ribeiro, A. (2023, 27 de dezembro). *Em primeiro ano de mandato, Lula repete alegações enganosas sobre fome, economia e governo Bolsonaro*. Aos Fatos. <https://www.aosfatos.org/noticias/lula-alegacoes-enganosas-2023/>
- Ribeiro, A. F. (1989). Correspondence in children's self-report: tacting and manding aspects. *Journal of the Experimental Analysis of Behavior*, 51, 361–367. <https://doi.org/10.1901/jeab.1989.51-361>

Skinner, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts.



WALDEN4 RESPONDE

A dúvida que chega ao consultório, respondida

Perguntas de quem procura terapia, de famílias que receberam um diagnóstico e de estudantes, respondidas por quem atende no dia a dia, com o que a pesquisa diz e o que ela ainda não sabe.

60+ respostas publicadas

Acesse, de graça e sem cadastro

walden4.com.br/responde



4

Atitudes políticas e preconceito como classes de equivalência

Muitas atitudes versam sobre coisas que a pessoa nunca experienciou. Guerin (1994) dá o exemplo da energia nuclear: o que a pessoa sabe sobre ela vem de jornais e de outras pessoas, não de contato com o evento. Em política o mesmo parece valer para boa parte do que se avalia, de partidos e instituições a grupos e a políticos que o eleitor só conhece por imagem e por nome. A pergunta que organiza este capítulo é como um nome, uma imagem ou um símbolo passa a ter função avaliativa para quem não teve contato com o que ele designa, e como essa função pode mudar.

Um experimento de 2021 mostra o fenômeno em miniatura. Cinco estudantes universitários aprenderam, por tentativa e erro, quais figuras abstratas “combinavam” com cada uma de três fotografias de presidentes do Brasil (Bolsonaro, Lula e Fernando Henrique Cardoso). Antes e depois do treino, avaliaram fotografias e figuras numa escala de adjetivos opostos. Três dos cinco avaliaram as figuras abstratas como neutras no pré-teste. No pós-teste, as figuras passaram a receber avaliações semelhantes às da fotografia a que tinham sido associadas (Pereira & Moreira, 2021b).

São cinco participantes, um delineamento de pré e pós-teste e uma escala. O capítulo trata o resultado como o que ele é: uma demonstração de que o mecanismo existe nessas condições, e não uma estimativa de quão comum ele é. O que o experimento oferece é um modelo explícito de como uma avaliação chega a um estímulo sem passar pelo contato com ele. As seções 4.1 a 4.3 constroem esse modelo; a 4.4 reúne os estudos com estímulos de significado social; a 4.5 descreve em detalhe os dois trabalhos com presidentes, o artigo de 2021b e os dois experimentos do livro de 2021a, com os estímulos e as telas que os próprios autores usaram; as seções 4.6 a 4.8 tratam das relações de oposição, da mudança de atitude e dos limites da evidência.

As figuras do capítulo seguem um código de cores que vale da primeira à última. Azul marca as fotografias dos presidentes (conjunto A); laranja e amarelo, as duas séries de figuras abstratas (B e C); roxo, as siglas dos partidos (D); verde, as palavras *direita*, *centro* e *esquerda* (E). Nas setas, o azul-escuro contínuo é a relação ensinada, com acerto e erro; o tracejado verde-escuro, a simetria testada; o tracejado roxo, a transitividade. Fotografias, figuras abstratas e telas vêm dos trabalhos de Pereira e Moreira (2021a, 2021b); os gráficos refazem os números das tabelas dos autores; os esquemas feitos para ensinar a ler um gráfico trazem escrito, na própria figura, que os valores são inventados.

4.1 Atitude como comportamento verbal avaliativo

Na psicologia social, o núcleo da definição é a avaliação. Ajzen (2001) registra que há acordo geral em que a atitude representa uma avaliação-síntese de um objeto psicológico, em dimensões como bom-mau e agradável-desagradável. Bohnert e Dickel (2011, p. 392) a definem como “uma avaliação de um objeto de pensamento” (tradução minha) e chamam de objeto de atitude tudo o que a pessoa pode ter em mente, do mundano ao abstrato: coisas, pessoas, grupos e ideias. Albarracín e Shavitt (2018) dizem que estudar atitudes é estudar avaliações. A definição já foi mais ampla, com componentes cognitivo, afetivo e comportamental, e foi se concentrando no componente avaliativo (Mizael et al., 2016b). A disputa que resta é sobre o estatuto da avaliação: Bohnert e Dickel (2011) ordenam as definições num eixo que vai de entidades estáveis, guardadas na memória, a julgamentos temporários, construídos na hora.

A Análise do Comportamento desloca a pergunta: em vez de que estado a avaliação revela, que contingências a mantêm. Para Guerin (1994), atitudes e crenças são comportamento verbal, e a modelagem por uma comunidade verbal é parte essencial da formação e da manutenção de ambas. Mizael et al. (2016b) resumem o que muda: para a área, atitudes não são apenas relacionadas a comportamentos, são comportamento, aberto ou encoberto, emitido ou evocado conforme uma história de aprendizagem, e, como comportamento verbal, podem assumir funções de tato, de mando e outras, conforme as contingências da comunidade. Isso permite investigar a formação e a mudança de atitudes a partir das relações com o ambiente, e não como algo causado por instâncias “dentro da pessoa” (p. 125). Cameoka e Moreira (2021) escrevem que, entendidas como comportamento verbal, independentemente da topografia, atitudes são comportamentos mantidos pelo reforço social generalizado da comunidade verbal a que o indivíduo pertence.

Ofereço, para os fins deste livro, uma definição em três camadas, cada uma acrescentando uma restrição à anterior. A primeira é a da psicologia social: atitude é a avaliação de um objeto, com um valor que vai do positivo ao negativo. A segunda é analítico-comportamental: essa avaliação é comportamento, selecionado e mantido por consequências, e não um estado que causa o comportamento. A terceira especifica o tipo: é comportamento verbal, mantido pelo reforço social que a comunidade verbal arranja, e por isso varia entre tato, intraverbal e mando conforme a função que cumpre para quem ouve (Capítulo 3). As duas primeiras camadas deixam o ouvinte de fora. A terceira o recoloca, e é por ela que se pode ler uma declaração de atitude como Guerin a leu, pelo efeito que tem em quem ouve (Capítulo 2, seção 2.8).

CONCEITO EM USO

Atitude (Guerin, 1994; Cameoka e Moreira, 2021)

Atitude é a resposta de atribuir a um objeto (coisa, pessoa, grupo ou ideia) um valor que vai do positivo ao negativo, quando essa resposta é verbal: mantida pelas consequências que a comunidade verbal arranja para ela e, por isso, capaz de funcionar como tato, intraverbal ou mando. O objeto não precisa ter sido contatado: o controle pode estar em outras palavras e imagens. A atitude difere da crença porque o falante pode sustentá-la diante de evidência contrária (Capítulo 2, seção 2.8).

Dois cuidados de uso. O primeiro é com a palavra *preconceito*, que aqui segue Pereira e Moreira (2021b): uma atitude em que os atributos avaliativos são negativos. Mizael e de Rose (2017) a descrevem como conjunto de atitudes negativas em relação a indivíduos com determinadas características físicas, cor ou etnia. O rótulo é do pesquisador e se aplica a uma relação (grupo-atributo negativo) que ele escolheu medir; o rótulo se mantém com essa restrição, como se mantém “discurso de ódio” como rótulo externo (Capítulo 5). O segundo cuidado é com a medida. O que os experimentos registram é uma marcação numa escala ou uma escolha numa tela, dirigida a um pesquisador. Isso é comportamento verbal sob controle de quem pergunta (Guerin, 1994), e a seção 4.8 volta ao que se pode e não se pode concluir dele.

4.2 Como uma palavra adquire função sem contato com o referente

Como um nome chega a ter valor para quem nunca contatou o que ele designa? Dixon et al. (2003) respondem com um exemplo escolhido por ser simples. Uma pessoa pode ouvir, numa ocasião, que os membros de um grupo nacional são grandes bebedores e, em outra, que grandes bebedores são um pouco

obtusos (no original, *dim-witted*). Ninguém lhe ensinou que os membros do grupo são obtusos, mas, segundo os autores, as duas relações ensinadas bastariam para derivar a relação entre o nome do grupo e o atributo; um preconceito infundado pode ser assim estabelecido por modos de discurso sobre o grupo, sem ensino direto. O artigo é conceitual: o exemplo é uma ilustração, não um dado.

Para saber se relações ensinadas produzem de fato relações não ensinadas, é preciso um procedimento em que as ensinadas sejam controladas e as derivadas, testadas. É o que oferece o paradigma de equivalência de estímulos, e é a partir dele que a literatura analítico-comportamental passou a entender a formação de atitudes como uma rede de relações arbitrárias entre classes de estímulos e atributos avaliativos (Mizael et al., 2016b). Dois conceitos bastam; apresento-os em quadros porque o capítulo os usa daqui em diante.

CONCEITO EM USO

Discriminação condicional e pareamento ao modelo (Sidman e Tailby, 1982)

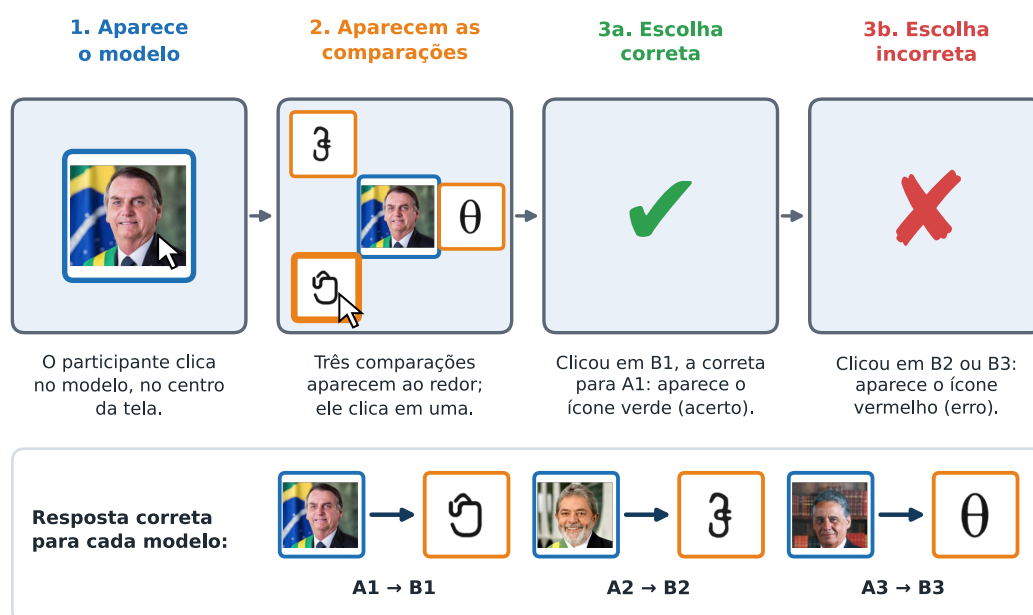
Numa tentativa de pareamento ao modelo (*matching-to-sample*, MTS), apresenta-se um estímulo modelo e dois ou mais estímulos de comparação, e a escolha de um deles é seguida de consequência de acerto; qual é o correto depende do modelo (Mizael et al., 2021b). O desempenho define uma relação condicional entre estímulos: se A1, então B1; se A2, então B2 (Sidman & Tailby, 1982). A Figura 4.1 mostra uma tentativa, com os estímulos de um dos experimentos que o capítulo descreve adiante.

A figura acompanha uma tentativa do experimento online descrito na seção 4.5. No primeiro quadro, o participante vê no centro da tela o estímulo modelo, aqui a fotografia de Bolsonaro, que os autores chamam de A1, e clica nele. No segundo, três figuras abstratas aparecem ao redor, e ele clica em uma. Qual delas é a correta não está escrito em lugar nenhum da tela: foi definido pelo pesquisador, e o participante só o descobre pelas consequências. Clicar em B1 diante de A1 é seguido de um ícone verde (terceiro quadro), e clicar em B2 ou em B3, de um ícone vermelho (quarto quadro); depois disso, a tela fica em branco por um segundo e começa outra tentativa (Pereira & Moreira, 2021a). A faixa de baixo resume o que se ensina: diante da fotografia de Lula (A2), a escolha correta é B2; diante da de Fernando Henrique Cardoso (A3), é B3. Correta, aqui, é só o que o procedimento reforça: nada na tarefa diz por que aquela figura combinaria com aquela fotografia. No artigo, a tela era um pouco diferente (o modelo no alto, as comparações embaixo, lado a lado, e um emoji feliz ou triste como consequência), mas a lógica da tentativa é a mesma (Pereira & Moreira, 2021b).

A condicionalidade, porém, não diz que os estímulos tenham outra relação além da que o procedimento ensinou. Sidman e Tailby (1982) observam que a equivalência, ao contrário da condicionalidade, não pode ser definida só pela interação do sujeito com o procedimento que a estabeleceu: exige testes, derivados das três propriedades que a matemática elementar especifica como definição da relação de equivalência.

FIGURA 4.1

Uma tentativa de pareamento ao modelo, passo a passo



Nota. Montada com as fotografias e as figuras abstratas dos dois trabalhos de Pereira e Moreira (2021a; 2021b) e com a disposição da tela do experimento online (2021a), reproduzida na Figura 4.10. Lê-se “A1→B1” como: com A1 como modelo, escolher B1.

CONCEITO EM USO

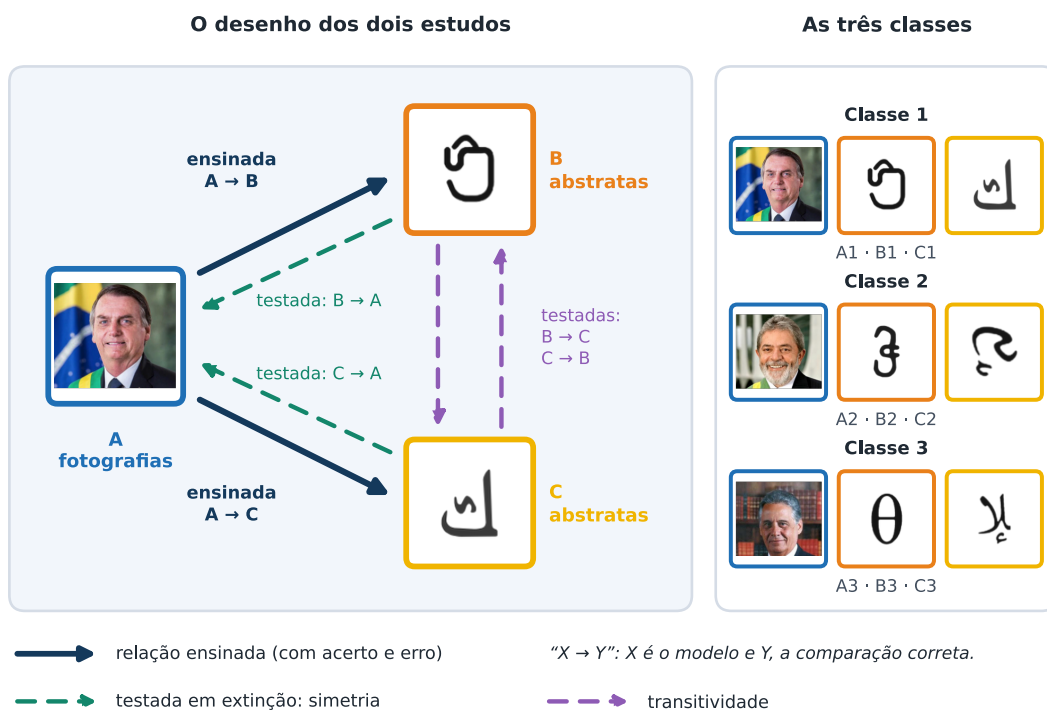
Classe de equivalência (Sidman e Tailby, 1982)

Um conjunto de estímulos forma uma classe de equivalência quando as relações condicionais entre eles têm as três propriedades da relação de equivalência. *Reflexividade*: cada estímulo se relaciona a ele mesmo ($A \rightarrow A$), o que se testa com pareamento por identidade generalizado. *Simetria*: ensinada $A \rightarrow B$, emerge $B \rightarrow A$ sem treino adicional. *Transitividade*: ensinadas $A \rightarrow B$ e $B \rightarrow C$, emerge $A \rightarrow C$ sem reforçamento diferencial. O teste que combina simetria e transitividade ($C \rightarrow A$, depois de $A \rightarrow B$ e $B \rightarrow C$) é o teste de equivalência (Sidman & Tailby, 1982; Mizael et al., 2016b). No paradigma original, com três conjuntos de três estímulos, seis relações ensinadas e seis emergentes definem três classes de três membros (Sidman & Tailby, 1982).

Os dois estudos com presidentes, descritos na seção 4.5, usam o paradigma com uma variação: em vez de uma cadeia $A \rightarrow B \rightarrow C$, ensinam-se duas relações a partir do mesmo estímulo, $A \rightarrow B$ e $A \rightarrow C$, em que A são fotografias de presidentes e B e C, formas abstratas (Figura 4.2). Nenhuma relação entre B e C é ensinada. As relações de B para A, de C para A e entre B e C são testadas sem consequências, e, quando as escolhas seguem o arranjo esperado, diz-se que as três classes, uma por fotografia, se formaram. As relações de B para C e de C para B exigem combinar a volta de B para A, que é simetria, com a ida de A para C; na terminologia do quadro acima, são testes de equivalência, embora Pereira e Moreira (2021b) as chamem de relações de transitividade.

FIGURA 4.2

Relações ensinadas, relações testadas e classes resultantes nos estudos com presidentes



Nota. Desenho dos estudos de Pereira e Moreira (2021a, 2021b), com as fotografias e as figuras abstratas dos próprios trabalhos (2021a; 2021b). Em cada tentativa de treino, o modelo é uma fotografia (A) e as comparações são as três figuras de um mesmo conjunto (B ou C). As setas tracejadas indicam relações que o procedimento não reforça: são testadas sem consequências. "X → Y" lê-se: X é o modelo e Y, a comparação correta. A reflexividade não aparece no desenho: costuma ser assumida, e não verificada, em participantes verbais (Mizael et al., 2016b).

Duas precisões evitam leituras erradas do que se chama classe. A primeira é que a classe se define por testes, não por semelhança: os membros não têm atributos físicos semelhantes e, depois do treino, tornam-se substituíveis uns pelos outros (Cameoka & Moreira, 2021), em certos contextos (Pereira & Moreira, 2021b). A segunda é que substituíveis não quer dizer equivalentes em

qualquer ocasião: tratamos a palavra “bolo” e o bolo como equivalentes e, ainda assim, não comemos a palavra (Mizael & de Rose, 2017). A classe é uma relação sob contexto, e é assim que o termo é usado a seguir.

O interesse do paradigma está no que ele explica. De Rose et al. (1988) enunciaram o problema: um comportamento novo, que ocorre sem história de reforçamento explícita e que não se explica por generalização primária de estímulos. Para estudá-lo, ensinaram a adultos, a crianças pré-escolares e a adultos com deficiência intelectual duas coisas. A primeira foi uma discriminação simples: com as formas A1 e A2 na tela ao mesmo tempo, escolher A1 era reforçado (A1 era o S+) e escolher A2, não (A2 era o S-). A segunda foi uma discriminação condicional por pareamento ao modelo: com A1 ou A2 como modelo e as formas B1 e B2 como comparações, escolher B1 diante de A1 e B2 diante de A2 era reforçado, de modo que B1 e B2 eram corretos com a mesma frequência. Depois, B1 e B2 foram apresentados juntos, sem modelo, e nenhuma escolha nessas sondas era reforçada. Pela história de reforçamento de B1 e B2, nada favorecia um deles, e mesmo assim os participantes quase sempre escolheram B1, a forma ligada condicionalmente a A1, o S+ da discriminação simples. Os autores sugerem que B1 e B2 adquiriram, no treino condicional, funções semelhantes às de A1 e A2. É o primeiro passo do que a seção seguinte chama de transferência de função.

4.3 Transferência de função

O que torna a classe de equivalência relevante para a atitude é a transferência de função: a função que um estímulo adquire por contato direto aparece também nos demais membros da classe, sem contato direto. Dougher et al. (1994) definem transferência de função como a aquisição de função de estímulo em virtude da pertinência a uma classe de equivalência. Dymond e Rehfeldt (2000) falam em transformação de funções quando as funções de um estímulo alteram as de outro de acordo com a relação derivada entre eles, sem treino adicional. A escolha do nome tem história: a maior parte da pesquisa foi relatada como transferência, e os autores defendem a terminologia genérica da teoria das molduras relacionais (p. 239); Sidman, por sua vez, objetou que “transferência” sugere um processo hipotético desnecessário, e Hayes e colegas, que o termo descreve mal o que acontece em relações que não são de equivalência (Weinstein et al., 2008). Uso *transferência de função* para o que ocorre por equivalência e *transformação* quando a relação é outra (seção 4.6).

Os experimentos que sustentam isso usam estímulos arbitrários e funções fáceis de medir. Funções discriminativas se estendem pelas relações condicionais, como no experimento de De Rose et al. (1988) descrito na seção anterior.

Funções de consequência também. Em Hayes et al. (1991), nove estudantes de psicologia introdutória aprenderam, por pareamento ao modelo com formas visuais arbitrárias, três classes de três membros. Depois, os estímulos de comparação receberam função de reforçador condicionado ou de punidor ao serem pareados com feedback verbal numa tarefa de classificação. Em oito dos nove participantes, a função treinada se transferiu pelas respectivas classes sem treino adicional.

CONCEITO EM USO

Transferência (ou transformação) de função (Dougher et al., 1994; Dymond e Rehfeldt, 2000)

Dado um estímulo que adquiriu uma função por contato direto (discriminativa, de consequência, eliciadora ou avaliativa), a função é testada nos demais membros da classe sem treino adicional (Dymond & Rehfeldt, 2000). Que a função tenha aparecido nos demais é dado; que tenha “transferido” pela classe é interpretação, a ser distinguida de explicações por contato direto (seção 4.8).

Funções eliciadoras também. Em Dougher et al. (1994), oito universitárias aprenderam, por pareamento ao modelo, duas classes de quatro membros. Um membro de uma classe foi pareado com choque elétrico, e um membro da outra classe foi apresentado sem choque; em seguida, todos os demais estímulos foram apresentados, e o condicionamento foi medido pela condutância da pele, uma medida fisiológica de resposta emocional. A resposta apareceu nos demais membros da classe do estímulo pareado com choque em seis dos oito participantes. No segundo experimento, depois de condicionar todos os membros de uma classe, um deles foi apresentado em extinção, isto é, sem choque, e os demais estímulos da classe deixaram de eliciar a resposta; quando o estímulo extinto foi recondicionado, os outros membros recuperaram a função eliciadora.

Um achado de Hayes et al. (1991) importa para a leitura de estudos com estímulos socialmente carregados. No Experimento 4, mesmo estímulos novos de feedback foram tratados como consequências: a própria estrutura da tarefa estabelece o feedback como consequência. A direção do efeito, porém, dependeu da transferência de funções específicas pelas classes. Em outras palavras, o procedimento já traz funções para dentro do laboratório, e é preciso separá-las das que a classe transferiu.

O elo com atitudes vem de estudos em que a função transferida é avaliativa. Em Bortoloti e de Rose (2009), universitários relacionaram, por pareamento ao modelo, faces com expressões emocionais a figuras arbitrárias e formaram três classes, cada uma com uma expressão emocional e três figuras arbitrá-

rias. Os julgamentos das figuras num diferencial semântico, uma escala de pares de adjetivos opostos, ficaram semelhantes às avaliações das faces, e mais ainda quando o treino usou pareamento com atraso, isto é, com um intervalo entre o modelo e as comparações. Num segundo experimento, figuras a um nó de distância das faces foram julgadas semelhantes a elas, e figuras a três nós, não. A distância nodal, isto é, quantos estímulos separam o membro avaliado do que tem a função, entra assim como variável: a transferência não é uniforme dentro da classe. Weinstein et al. (2008) levaram a ideia ao estigma. Cinquenta universitários (45 analisados) aprenderam por pareamento ao modelo a responder a formas arbitrárias como se fossem estímulos ligados a obesidade ou a magreza. Depois, num teste de associação implícita (IAT), que estima o viés pela rapidez com que a pessoa associa categorias a palavras positivas ou negativas, com formas semelhantes às treinadas, apareceu um efeito de viés pequeno ($d = 0,31$), contra $d = 0,96$ no IAT tradicional de obesidade. O d é uma medida de tamanho de efeito, a diferença entre médias em unidades de desvio-padrão; por convenção, 0,2 é pequeno e 0,8, grande. Os autores atribuem a diferença à duração: cerca de 45 minutos de treino contra uma vida de história.

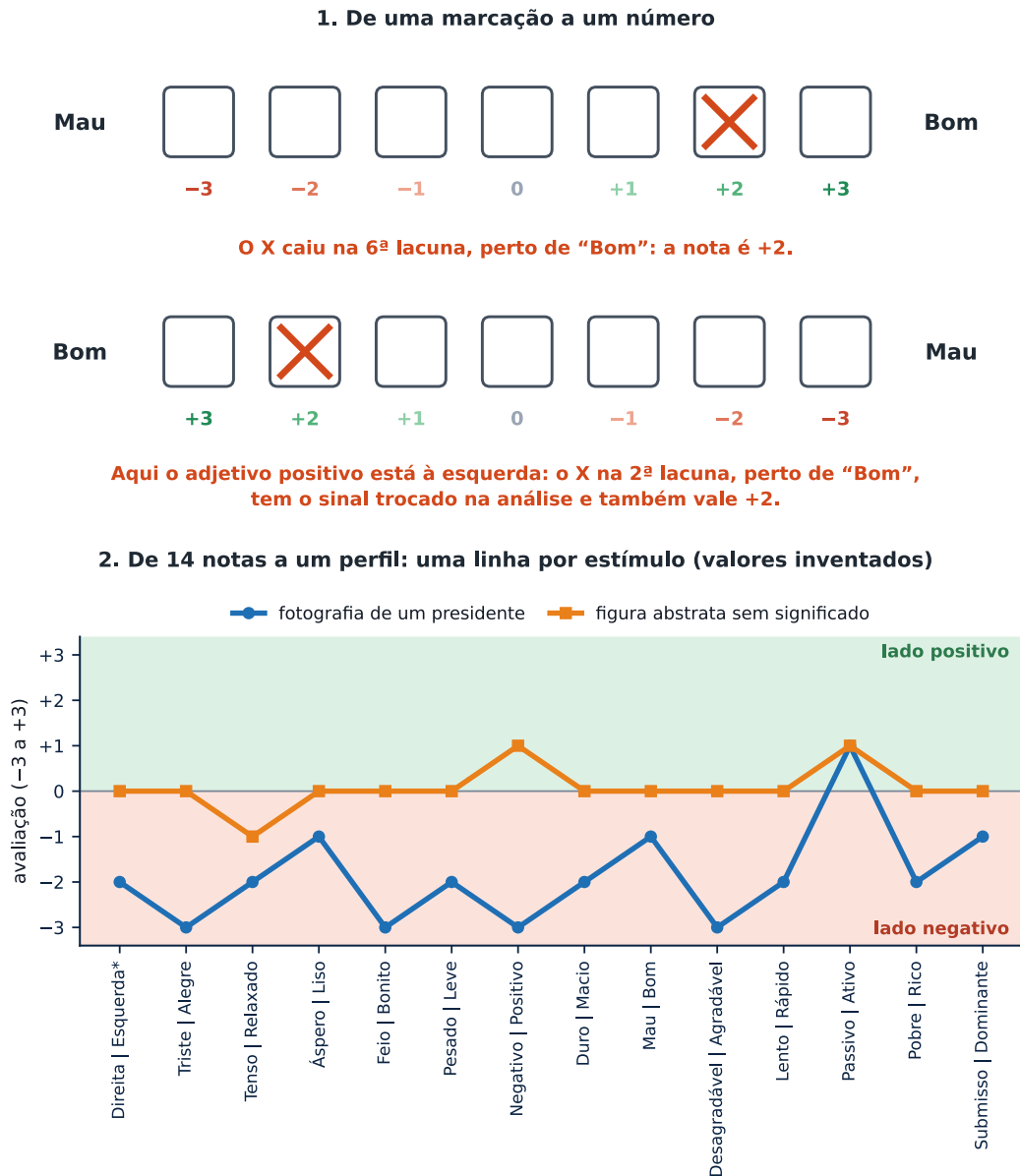
Como se lê uma escala de diferencial semântico

Os estudos com presidentes, que a seção 4.5 descreve, medem a avaliação com uma escala de diferencial semântico, e saber ler os seus gráficos é o que permite acompanhar os resultados. A escala de Pereira e Moreira (2021b) tinha 14 pares de adjetivos bipolares, como mau e bom, com sete lacunas entre os adjetivos de cada par. Cada folha trazia um estímulo, e o participante marcava, em cada par, um X na lacuna que lhe parecia descrever melhor o estímulo: a mais próxima de um adjetivo, se o descrevia muito; a do meio, se nenhum dos dois. Em metade dos pares o adjetivo positivo ficava à esquerda e na outra metade, à direita, de modo que marcar sempre do mesmo lado não produz uma avaliação uniforme.

A Figura 4.3 mostra, na parte de cima, como uma marcação vira um número. Para a análise, os autores deram a cada lacuna um valor de -3 a +3 e trocaram o sinal dos pares em que o adjetivo positivo estava à esquerda, de modo que, em todos os pares, o adjetivo negativo ficou do lado de -3 e o positivo, do lado de +3. Na parte de baixo da figura, as 14 notas de um estímulo, ligadas por uma linha, formam o perfil de avaliação desse estímulo. Lê-se de quanto a linha se afasta do zero, para o lado negativo ou para o positivo, e comparar duas linhas é comparar duas avaliações. Os autores escolheram os gráficos de linhas porque facilitam ver a “forma da avaliação” de cada estímulo e compará-las, ainda que as notas de pares diferentes não sejam uma medida contínua de um mesmo comportamento.

FIGURA 4.3

Como se lê uma escala de diferencial semântico: de uma marcação a um número e de 14 números a um perfil



* Direita/Esquerda: o artigo converte esse par com a convenção própria descrita na seção 4.8.

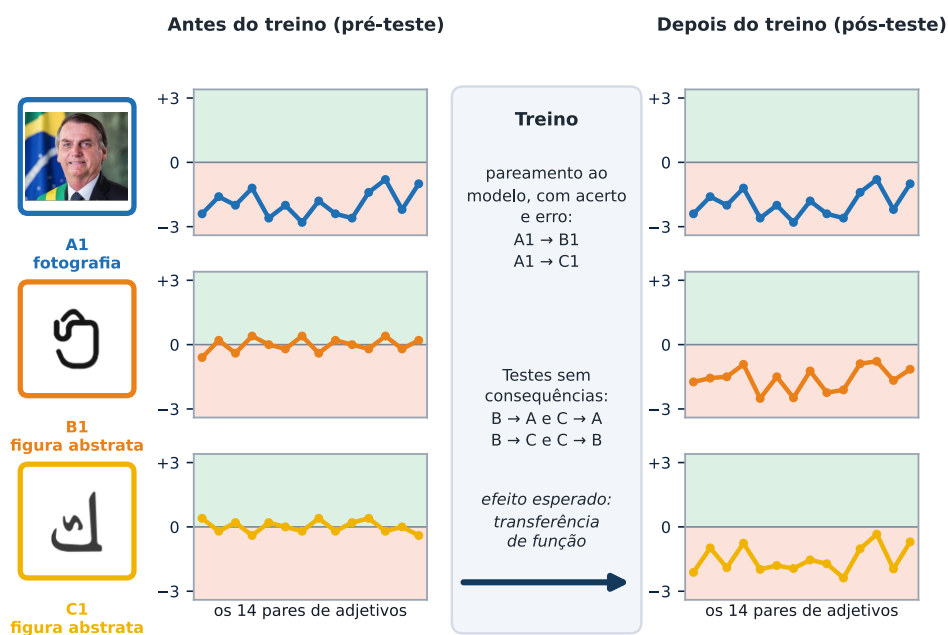
Nota. Esquema do autor. Os 14 pares são os da escala de Pereira e Moreira (2021b), com o adjetivo negativo à esquerda, como na análise; no par Submisso-Dominante, o artigo não diz qual adjetivo foi tratado como negativo, e a figura segue a ordem da lista do artigo. Os valores dos dois perfis são inventados, só para ensinar a ler o gráfico. * O par Direita-Esquerda segue convenção própria do artigo (seção 4.8).

Como seria a transferência de uma avaliação

A Figura 4.4 junta as peças num esquema do que a transferência de função quer dizer para uma avaliação. À esquerda, os três estímulos de uma classe, tirados do experimento de Pereira e Moreira: a fotografia de um presidente (A1) e duas figuras abstratas (B1 e C1). No pré-teste, a fotografia tem uma avaliação própria, negativa neste exemplo, e as figuras abstratas, que o participante nunca viu, ficam perto do zero, a marcação neutra, que, como a seção 4.5 lembra, não é prova de ausência de significado. Ao centro está o treino, em que A1 é modelo para B1 e para C1, com acerto e erro, seguido de testes sem consequências, que verificam se as figuras remetem de volta à fotografia e se se relacionam entre si. À direita, o que se espera ver se a função foi transferida: as duas figuras passam a ter avaliações parecidas com as da fotografia, que não mudou.

FIGURA 4.4

Transferência de função de uma avaliação por uma classe de equivalência



Valores inventados, só para mostrar a ideia. Cada ponto é um par de adjetivos (−3 a +3); a faixa rosa é o lado negativo e a verde, o positivo.

Nota. Esquema do autor, com valores inventados, só para mostrar a ideia; não é dado de nenhum participante. Os estímulos são os do experimento de Pereira e Moreira (2021b), cujo treino e testes a seção 4.5 descreve. Nos experimentos citados nesta seção com outras funções, o evento é um choque (Dougher et al., 1994) ou feedback verbal (Hayes et al., 1991), e os estímulos são formas arbitrárias.

■ DADO	Em experimentos com estímulos arbitrários, funções discriminativas, de consequência, eliciadoras e avaliativas observadas em um membro de uma classe de equivalência apareceram em outros membros sem treino direto (De Rose et al., 1988; Hayes et al., 1991; Dougher et al., 1994; Bortoloti & de Rose, 2009).
■ INTERPRETAÇÃO	Nessas condições, a pertinência à classe basta para que a função apareça em membros que não a tiveram por contato direto. A transferência pode variar com parâmetros do procedimento e com a distância nodal.
■ HIPÓTESE	Que o mesmo processo opere sobre nomes, imagens e símbolos de grupos e de políticos, na história de uma pessoa fora do laboratório, é extrapolação. As seções 4.4 e 4.5 examinam versões dela, com os limites ali descritos.

4.4 Preconceito nos experimentos

O preconceito entra nos experimentos de equivalência por uma diferença de material. Com formas arbitrárias, o participante chega ao laboratório sem história com os estímulos; com faces, nomes, emblemas e palavras que designam grupos, não: ele chega com relações já mantidas pela comunidade verbal. Os estudos desta seção se dividem, grosso modo, em dois desenhos. Num, o treino ensina relações que não conflitam com essa história, e se pergunta se a função se transfere para estímulos arbitrários (Weinstein et al., 2008; Pereira & Moreira, 2021b). No outro, o treino ensina relações que a contrariam, e se pergunta se a classe se forma e se a atitude muda. Mizael et al. (2016a) deram a este segundo desenho o nome de *Conflicting Relations Paradigm* (Mizael et al., 2016b). A Tabela 4.1 reúne os estudos com estímulos de significado social cujo texto integral foi lido; os dois trabalhos com presidentes, que fazem parte do primeiro desenho, são descritos à parte, na seção 4.5.

TABELA 4.1

Experimentos de equivalência com estímulos de significado social: o que se fez, o que se observou e o que não permitem dizer

Estudo	Participantes e estímulos	O que se observou	Ressalva
Weinstein et al. (2008)	50 universitários (45 analisados); formas arbitrárias relacionadas, por pareamento ao modelo, a palavras sobre obesidade e magreza; medida por IAT	Viés no IAT com formas semelhantes após o treino ($d = 0,31$), contra $d = 0,96$ no IAT tradicional	Treino de cerca de 45 min; cinco excluídos por baixa fluência; autorrelato não se correlacionou com o IAT tradicional
Dixon e Lemke (2007)	12 universitários norte-americanos; fotografias de homens do Oriente Médio, de homens americanos e de objetos; escala de 1 (mau) a 10 (bom)	Maioria inverteu a ordem das avaliações do pré-teste: homens do Oriente Médio passaram a receber avaliações mais altas que os americanos	Amostra pequena; o viés se ligava, provavelmente, aos atentados de 2001, mais intergrupais que racial (Cameoka & Moreira, 2021)
McGlinchey e Keenan (1997)	36 participantes na Irlanda do Norte (28 crianças de 5 a 11 anos, dois jovens de 12 e seis adultos); classes com bandeira britânica, estímulo arbitrário e dançarina irlandesa	23 das 28 crianças responderam por equivalência (as outras cinco, não em ao menos um teste); os dois jovens de 12 anos e os seis adultos, em nenhum (13 de 36, ou 36%)	A idade confunde duração da história com outras diferenças (ressalva do autor)
Haydu et al. (2015)	28 homens de torcidas organizadas de três clubes paulistas, moradores de Londrina (PR); emblemas dos clubes, pinturas abstratas e as palavras Bom, Ruim e Regular	Relações de linha de base e simetria se formaram; transitividade e equivalência não emergiram; erros, em maioria, no sentido da história	Só homens, de uma cidade; palavras distribuídas de propósito para conflitar com a história
Moxon et al. (1993)	19 universitários (10 homens e 9 mulheres); ocupações tradicionalmente masculinas, sílabas sem sentido e nomes femininos; ocupações tradicionalmente femininas como estímulos novos no teste	Respostas de equivalência em 76% das possíveis nas mulheres e 52% nos homens; estímulo novo escolhido em 18% e 42% das ocasiões; generalização inconsistente, sem diferença entre os grupos	N = 19; estereotipia dos participantes não medida (leitura do autor): a explicação pela história é dos autores do estudo

Estudo	Participantes e estímulos	O que se observou	Ressalva
de Carvalho e de Rose (2014)	4 crianças de 7 a 10 anos com viés negativo contra homens negros, selecionadas entre 16	Uma reverteu as relações do pré-teste; para três, elas persistiram	N = 4; faces podem controlar o responder por expressão e outras propriedades (Cameoka & Moreira, 2021)
Mizael et al. (2016a)	13 crianças com viés negativo contra faces negras; formas relacionadas a faces negras e a um símbolo positivo; treino modificado	As 13 formaram as classes; nove mantiveram as relações num teste mais rigoroso; a diferença entre avaliações de faces negras e brancas deixou de ser significativa	Quatro de 13 não mantiveram no teste rigoroso (cálculo do autor)
Mizael et al. (2021b)	46 crianças com viés negativo, em quatro grupos (um controle e três com parâmetros diferentes); seguimento de seis semanas	33 formaram classes entre faces negras e símbolo positivo; 27 mantiveram ao menos uma relação em seis semanas; viés médio positivo no pós-teste para quem formou e negativo para quem não formou	Associação, não causa; o grupo controle teve redução de viés e manutenção menos pronunciadas

Nota. Os números vêm dos artigos (Weinstein et al., 2008; McGlinchey & Keenan, 1997; Haydu et al., 2015; Moxon et al., 1993; de Carvalho & de Rose, 2014; Mizael et al., 2016a; Mizael et al., 2021b; Mizael et al., 2016b, para Dixon & Lemke, 2007). A contagem 13 de 36 em McGlinchey e Keenan (1997) é do autor e confere com os “36% dos casos” do texto.

Três desenhos da tabela e de seus arredores mostram como esses estudos são montados. Em McGlinchey e Keenan (1997), seis adultos e trinta crianças, católicos e protestantes da Irlanda do Norte, aprenderam quatro discriminações condicionais: diante de A1, um estímulo protestante (a bandeira britânica), escolher B1, uma forma arbitrária; diante de A2, um estímulo católico (a dançarina irlandesa), escolher B2; e, depois, diante de B1 escolher C1, um estímulo católico, e diante de B2, C2, um protestante. Um teste de equivalência verificou se os símbolos católico e protestante tinham ficado relacionados por simetria e transitividade, e testes com estímulos novos verificaram se relações aprendidas socialmente competiam com a emergência da equivalência. Barnes et al. (1996) expuseram 12 crianças com deficiência intelectual leve e 12 sem deficiência a dois procedimentos de equivalência, um com formas abstratas, que servia de controle, e outro com as palavras “Slow” (lento) e “Able” (capaz) e com nomes pessoais, o da criança e um nome fictício, de modo que o treino ligava o nome da criança a “Able”. As crianças com deficiência tiveram menos respostas de equivalência nesse teste, isto é, falharam em ligar o próprio nome a “Able”, e os autores sugerem que relações pré-experimentais en-

tre o nome delas e o termo “Slow” podem ter perturbado a formação das relações induzidas. Em Merwin e Wilson (2005), 32 participantes responderam antes a duas medidas de autorrelato, de sofrimento e de autoestima, e foram divididos em grupos de sofrimento e de autoestima altos ou baixos; depois fizeram duas tarefas de pareamento ao modelo, uma que relacionava estímulos autorreferentes (eu, mim) a palavras de avaliação positiva e outra que os relacionava a palavras de avaliação negativa. Quem relatava sofrimento alto e autoestima baixa errou mais nos testes de equivalência da tarefa com palavras positivas.

A tabela não mostra um resultado; mostra um padrão com três partes. Primeira: com estímulos arbitrários, funções avaliativas aparecem depois de treinos curtos, em tamanho pequeno (Weinstein et al., 2008; Bortoloti & de Rose, 2009). Segunda: quando a classe ensinada contraria a história social, a formação é interrompida com frequência, e a interrupção acompanha a força da história: em McGlinchey e Keenan (1997), os indivíduos com histórias de aprendizagem social mais extremas, contrárias ao treino, foram os menos propensos a responder por equivalência. Em Moxon et al. (1993), com ocupações e nomes femininos, os homens escolheram o estímulo novo mais que as mulheres (42% e 18% das ocasiões), e os autores atribuem a diferença ao maior controle, entre os homens, de classes estereotípicas aprendidas socialmente; o estudo não mediu essa história. Barnes et al. (1996) e Merwin e Wilson (2005), descritos acima, apontam na mesma direção, com relações pré-experimentais entre o próprio nome e o termo *lento* e com o sofrimento e a baixa autoestima relatados. Terceira: o mesmo conteúdo social produz resultados opostos conforme o procedimento. De 1 em 4 crianças com viés negativo contra homens negros que reverteram as relações do pré-teste (de Carvalho & de Rose, 2014) a 13 em 13 que formaram as classes (Mizael et al., 2016a), com o mesmo tipo de estímulo; Mizael et al. (2016b) atribuem a diferença, ao menos em parte, a parâmetros de treino, como ir do protocolo simples ao complexo e reduzir gradualmente a consequenciação na última fase. A dificuldade, portanto, não é só da história: é também do desenho.

A Tabela 4.1 não esgota o que se fez no Brasil com esse desenho. O Instituto Walden4 reuniu, num livro de acesso livre, experimentos do mesmo tipo com estímulos de significado social: *Preconceito: doze experimentos e um paradigma* (Moreira et al., 2022b). Depois de um primeiro experimento sobre controle de estímulos e mensuração de atitudes, que a seção 4.8 retoma, o volume traz experimentos de transferência de função e de equivalência com estímulos relacionados ao vestuário da mulher, a gênero e esporte, ao Candomblé, à tatuagem em profissionais da área médica, à obesidade, à ressocialização de presos, a idosos no mercado de trabalho, a usuários de drogas e a biótipos corpo-

rais (sumário do volume). Eles ficam fora da tabela porque este capítulo não os leu em detalhe; para quem quer o que a tabela não mostra, são o ponto de partida mais próximo.

■ DADO	Com estímulos de significado social, a classe pretendida deixou de se formar em parte dos participantes (McGlinchey & Keenan, 1997: 13 de 36; Haydu et al., 2015: transitividade e equivalência não emergiram; de Carvalho & de Rose, 2014: 3 de 4 não revertiram as relações do pré-teste) e se formou em outros, com treinos modificados (Mizael et al., 2016a: 13 de 13; Mizael et al., 2021b: 33 de 46).
■ INTERPRETAÇÃO	A formação de classes que envolvem estímulos socialmente carregados depende da relação entre o que se ensina e a história do participante, e de parâmetros do treino. Onde a classe ensinada contraria a história, ela compete com relações já mantidas pela comunidade verbal.
■ HIPÓTESE	A competição cresce com a duração e a força da história de reforço, o que explicaria a diferença entre crianças pequenas e adultos em McGlinchey e Keenan (1997). Os estudos comparam idades (McGlinchey & Keenan, 1997) ou sexos (Moxon et al., 1993); nenhum manipula a duração da história.

4.5 Atitudes em relação a presidentes do Brasil

Dois trabalhos de Pereira e Moreira (2021a, 2021b) são, entre os estudos de equivalência verificados, os únicos com estímulos de política partidária; os autores dizem não conhecer outro (2021b). Um é o artigo da *Acta Comportamentalia* (2021b), com cinco estudantes de Psicologia, escala em papel e treino num computador da sala de aula. O outro é o livro (2021a), com dois experimentos de cinco estudantes cada, feitos de casa, pela internet. Esta seção descreve os dois com o detalhe de um relatório de método, um de cada vez, na ordem em que o participante os viveu: o que viu, o que fez e o que os números mostram. As Figuras 4.5 a 4.13 acompanham o texto.

O artigo: pergunta, participantes e delineamento

A pergunta do artigo era se um procedimento de pareamento ao modelo poderia induzir a transferência de função de estímulos com significado político, as fotografias dos presidentes, para estímulos sem significado, as figuras abstratas. Os autores esperavam que uma resposta afirmativa desse suporte ao paradigma de equivalência para o estudo de atitudes e acrescentasse evidências com estímulos de significado social que não se ligam a preconceito. A variável

independente foi o treino de discriminações condicionais; as variáveis dependentes foram a formação das classes, medida pela porcentagem de acertos nos testes, e a transferência de função, medida pela diferença entre as avaliações das figuras abstratas nas duas aplicações da escala. O delineamento é de sujeito único, com pré-teste e pós-teste: cada participante serve de comparação para si mesmo.

Participaram cinco estudantes universitários de 20 a 25 anos, do quinto ao sétimo semestre de Psicologia, escolhidos por conveniência, sem contato prévio com a tarefa; todos assinaram o termo de consentimento, e o projeto foi aprovado por um comitê de ética em pesquisa. A coleta foi numa sala de aula de uma instituição de ensino superior, com um computador conectado à internet, uma mesa e uma cadeira; o pareamento rodava num software online elaborado pelo segundo autor, que apresentava os estímulos e registrava as respostas. Cada sessão durou de 80 a 100 minutos.

Os estímulos e as relações ensinadas

Os estímulos formam cinco conjuntos (Figura 4.5). O conjunto A reúne as fotografias dos três presidentes: Bolsonaro (A1), Lula (A2) e Fernando Henrique Cardoso (A3). Os conjuntos B e C são duas séries de três figuras abstratas, sem significado conhecido para os participantes; o D tem as siglas dos partidos dos presidentes, PSL, PT e PSDB; e o E, as palavras *direita*, *centro* e *esquerda*. O que o participante aprende é a relacionar A com B e A com C. Cada fotografia tem, assim, uma classe pretendida, formada por ela e por uma figura de cada série: A1, B1 e C1; A2, B2 e C2; A3, B3 e C3. As siglas e as palavras não foram ensinadas: entram apenas no segundo teste, para verificar se o participante já as relacionava às fotografias e às figuras. Só os conjuntos A, B, C e D foram avaliados na escala, num total de 12 estímulos.

FIGURA 4.5*Os estímulos do artigo, por conjunto e por classe*

	Classe 1 Bolsonaro PSL	Classe 2 Lula PT	Classe 3 Fernando Henrique Cardoso · PSDB
A · fotografias dos presidentes modelo nos treinos; avaliadas na escala			
B · formas abstratas comparações nos treinos; avaliadas			
C · formas abstratas comparações nos treinos; avaliadas			
D · siglas dos partidos só no Teste 2; avaliadas na escala	PSL	PT	PSDB
E · palavras só no Teste 2; não avaliadas na escala	direita · centro · esquerda		

Nota. Fotografias e figuras abstratas reproduzidas de Pereira e Moreira (2021a; 2021b). As siglas e as palavras foram compostas para a figura. O artigo não diz qual palavra corresponde a qual coluna (E1, E2, E3), e por isso o conjunto E ocupa as três colunas.

As sete etapas

O procedimento teve sete etapas, em sessão única, com intervalos de cerca de três minutos entre elas, o tempo que a programação de cada etapa levava para ser feita no computador (Figura 4.6). No pré-teste, o pesquisador explicou oralmente como preencher a escala e entregou ao participante um bloco de 12 folhas de papel A4, uma para cada estímulo, avaliado uma única vez, numa ordem fixa que começava por A1, C3, D2 e B1; em cada folha, o participante marcava um X na escala de sete pontos em todos os 14 pares de adjetivos.

Os três treinos usaram pareamento ao modelo simultâneo. Em cada tentativa, o modelo, uma fotografia, aparecia na parte central superior da tela; o clique nele fazia surgirem três comparações na parte inferior, lado a lado, e o modelo permanecia na tela. Depois do clique numa comparação, todos os estímulos saíam da tela e aparecia, por um segundo, um emoji feliz, se a escolha era a programada como correta, ou triste, se não era, seguido de dois segundos de tela em branco. O Treino AB tinha 10 blocos de 18 tentativas, com cada uma

das três relações (A1 diante de B1, A2 diante de B2 e A3 diante de B3) apresentada seis vezes por bloco, em ordem aleatória; o critério era 100% de acerto em um bloco, atingido em até 30 minutos, e quem não o atingisse refaria o treino em outro dia. O Treino AC repetiu o AB com as figuras C, e o Treino AB AC misturou as seis relações, em 10 blocos de 18 tentativas, com três apresentações de cada uma.

FIGURA 4.6

As sete etapas do procedimento do artigo



Nota. Esquema do autor, a partir do método de Pereira e Moreira (2021b). A sessão inteira durou de 80 a 100 minutos.

Nos testes, ninguém dizia ao participante se a escolha era correta ou errada: o primeiro teste de emergência de relações não trazia consequências diferenciadas. Tinha quatro blocos de 18 tentativas, com seis apresentações de cada relação testada, e verificava, nesta ordem, as relações de simetria BA e CA (a figura como modelo e a fotografia como comparação) e as de transitividade BC e CB (uma figura como modelo e as figuras do outro conjunto como comparações). O critério era de 88,88% de acerto em cada bloco, isto é, 16 acertos em 18 tentativas, para afirmar que a relação tinha emergido. O segundo teste, de seis blocos, verificou relações não treinadas com as siglas e as palavras: AD, BD e CD, de um lado, e AE, BE e CE, do outro. Os autores ressaltam que o ter-

mo emergência não se aplica às relações das fotografias com as siglas e com as palavras (AD e AE), que provavelmente já existiam antes do experimento. O pós-teste foi idêntico ao pré-teste.

Como se calculam o Grau de Equivalência e o Grau de Transferência

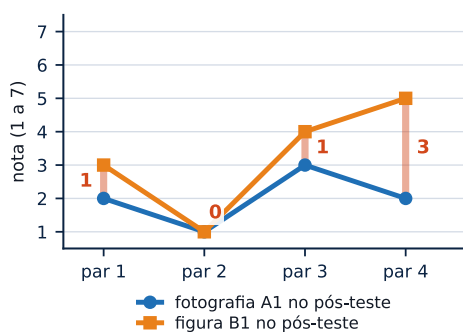
Para ir além da inspeção dos gráficos, os autores criaram duas medidas percentuais. O Grau de Equivalência (GE) mede a sobreposição entre as avaliações de dois estímulos diferentes de uma mesma classe num mesmo teste, por exemplo A1 e B1 no pós-teste; GE de 100% quer dizer avaliações idênticas nos 14 pares. O Grau de Transferência (GT) mede a diferença entre as avaliações de um mesmo estímulo em testes diferentes, por exemplo B1 no pré-teste e B1 no pós-teste; GT de 100% quer dizer que a avaliação mudou por completo, por exemplo de -3 para +3 nos 14 pares. A conta é a mesma para as duas (Figura 4.7). A escala é convertida de -3 a +3 para 1 a 7; toma-se o valor absoluto da diferença entre as avaliações dos dois estímulos em cada par; somam-se as diferenças dos 14 pares; como a maior diferença possível por par é 6, a maior soma é $6 \times 14 = 84$; o percentual dessa soma em relação a 84 é o GT, e o GE é 100 menos esse percentual.

FIGURA 4.7

Como se calculam o Grau de Equivalência (GE) e o Grau de Transferência (GT)

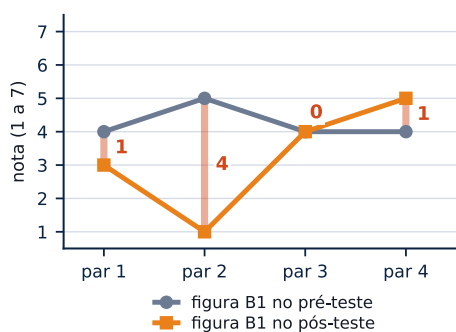
A mesma conta, para duas perguntas (exemplo inventado, com 4 pares de adjetivos)

GE: dois estímulos, no mesmo momento



1. Diferença em cada par: 1, 0, 1, 3 (soma 5)
2. Máximo possível: $6 \times 4 \text{ pares} = 24$
3. $5 \div 24 = 20,8\%$ de diferença
4. $GE = 100 - 20,8 = 79,2\%$

GT: o mesmo estímulo, em dois momentos



1. Diferença em cada par: 1, 4, 0, 1 (soma 6)
2. Máximo possível: $6 \times 4 \text{ pares} = 24$
3. $6 \div 24 = 25,0\%$
4. $GT = 25,0\%$ (sem subtrair de 100)

Com os 14 pares do estudo, o máximo é $6 \times 14 = 84$. $GE = 100\%$ quer dizer avaliações idênticas; $GT = 100\%$, mudança total.

Nota. Exemplo do autor, com quatro pares de adjetivos e valores inventados, para mostrar a conta; o estudo usou 14 pares, e o máximo da soma foi 84 (Pereira & Moreira, 2021b).

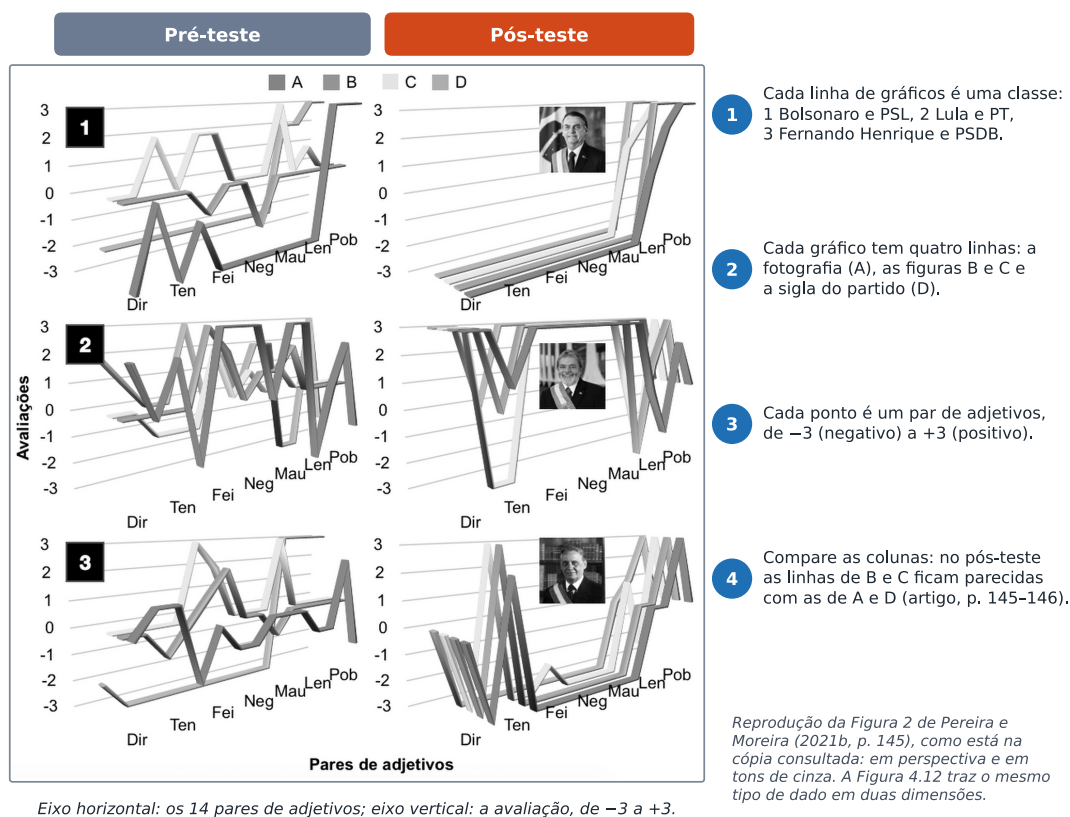
Os autores esperavam, com isso, que estímulos com significado pré-experimental, como as fotografias, tivessem GT próximo de 0%, isto é, que suas avaliações não mudassem; que estímulos sem significado, como as figuras abstratas, tivessem GT próximo de 50%; e que esses mesmos estímulos tivessem GE próximo de 100% com as fotografias das suas classes. Um GT de 50% equivale, em média, a mover três pontos em cada par da escala de seis pontos de amplitude, o que corresponde, por exemplo, a ir do centro da escala a um dos extremos; a conta é do autor deste livro, e não aparece no artigo.

O que o artigo encontrou

Todos os participantes atingiram o critério de 100% de acerto nos treinos (AB, AC e AB AC), na primeira execução, e, no primeiro teste, todos atingiram o critério de 88,88%, exceto P02 na relação BA (72,22%); com essa ressalva, os autores concluem que se formaram as três classes programadas. No segundo teste, todos acertaram 94,44% ou mais. Na Tabela 1 do artigo, quase todas as células são 100%: os demais valores são 94,44% (P01 nas relações CB e BE; P05 nas CB e BD) e, em P02, 72,22% em BA e 94,44% em CA.

A Figura 4.8 reproduz a Figura 2 do artigo, com os gráficos de P01. A coluna da esquerda é o pré-teste e a da direita, o pós-teste; cada linha de gráficos é uma classe; em cada gráfico há quatro linhas, a da fotografia, as das figuras B e C e a da sigla do partido. Para P01, na classe 1, a fotografia de Bolsonaro e a sigla PSL tiveram avaliações predominantemente negativas no pré-teste, com exceção dos três últimos pares; a figura B1 ficou perto de -1 e a C1, entre zero e três. No pós-teste, a fotografia e a sigla mantiveram as avaliações, e B1 e C1, diferentes das do pré-teste, ficaram parecidas com as da fotografia. Na classe 2, Lula e PT foram avaliados de modo predominantemente positivo no pré-teste, e, no pós-teste, as figuras B2 e C2 passaram a acompanhar a fotografia. A classe 3 mostra outro padrão: no pré-teste, a fotografia de Fernando Henrique Cardoso variou entre neutra e negativa, e a sigla PSDB foi predominantemente negativa; no pós-teste, a sigla manteve a avaliação, e a fotografia e as figuras ficaram parecidas com ela, o que os autores leem como transferência da sigla para os demais estímulos da classe.

Os resultados dos outros quatro participantes foram semelhantes aos de P01 quanto à transferência de função. Varia, entre eles, o ponto de partida das figuras abstratas: P02, mais que P01, deu avaliações positivas e negativas a elas já no pré-teste, o que, a rigor, sugere que as figuras tinham significado pré-experimental para ele, e P03, P04 e P05 deram nota zero, neutra.

FIGURA 4.8*Avaliações de P01 no pré e no pós-teste, nas três classes*

Nota. Reproduzida de Pereira e Moreira (2021b), como está na cópia consultada: em perspectiva e em tons de cinza, embora o artigo diga que os gráficos foram feitos em cores. Foram acrescentados as faixas de pré e pós-teste e o guia de leitura. As figuras equivalentes dos outros participantes são as Figuras 3 a 6 do artigo.

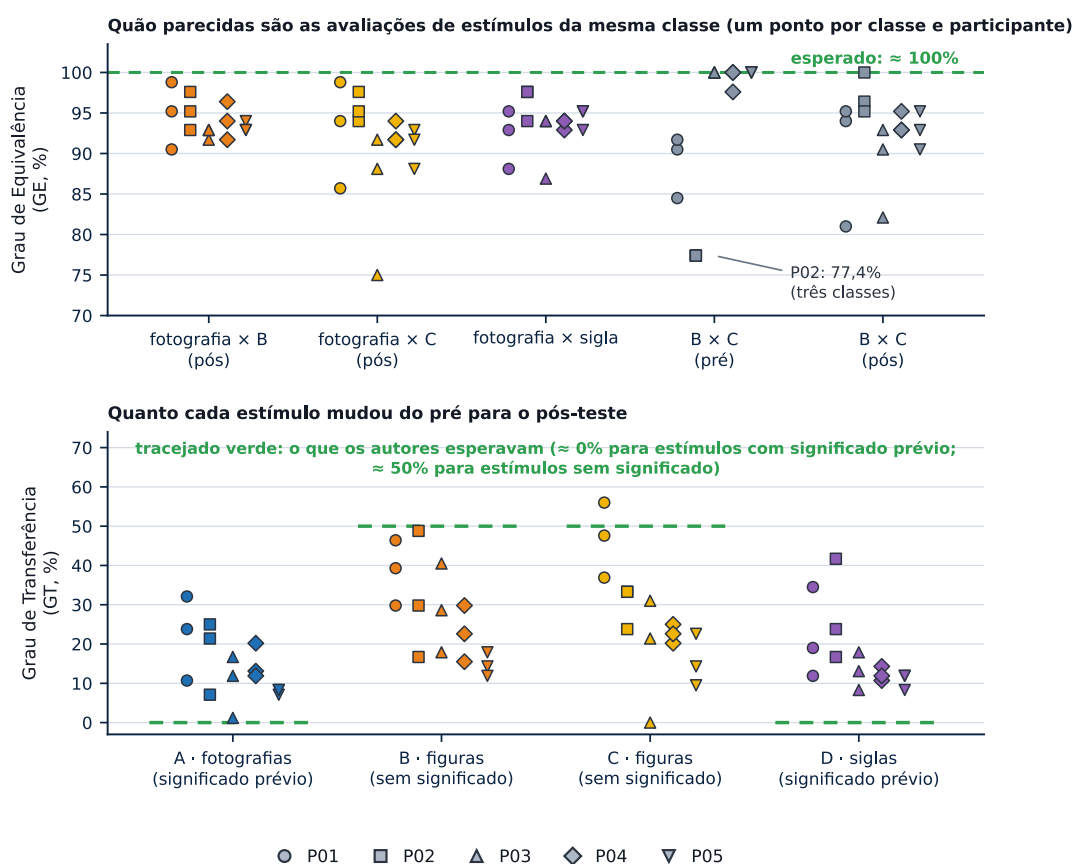
A Figura 4.9 reúne os números da Tabela 2 do artigo, para os cinco participantes e as três classes: cada coluna tem 15 pontos. Na parte de cima, o GE no pós-teste entre cada fotografia e a figura B da sua classe ficou acima de 90% em todos os participantes e classes; entre fotografia e figura C, o padrão se repetiu, com diferenças inferiores a 5 pontos em relação ao de B, exceto na classe 3 de P03 (75,0%, contra 92,9% na figura B). O GE entre as figuras B e C, no pré-teste, foi maior que 90% na maioria dos casos, exceto em P02, em que ficou em 77,4% nas três classes. O GE, sozinho, poderia induzir a uma leitura enganosa, se as avaliações de B no pré-teste já fossem parecidas com as de A, e por isso foram calculados também os GT.

Na parte de baixo da figura, o GT das figuras B foi de 11,9% (P05, B3) a 48,8% (P02, B1); o das figuras C, na tabela, vai de 0,0% (P03, classe 3) a 56,0% (P01, classe 1), e só esse valor passa dos 50% que os autores esperavam para estímulo-

los sem significado. As fotografias e as siglas, que se esperava estáveis, tiveram GT de 1,2% a 32,1% e de 8,3% a 41,7%, respectivamente. Para P01, por exemplo, a avaliação da fotografia de Lula variou 23,8% e a de Fernando Henrique Cardoso, 32,1%. Os autores admitem que não fica claro se essas variações nos estímulos com significado prévio são uma margem de erro de medidas repetidas ou transferência de função dos demais estímulos para as fotografias, e especulam que a transferência talvez não seja um fenômeno unidirecional.

FIGURA 4.9

Grau de Equivalência e Grau de Transferência dos cinco participantes



Nota. Valores da Tabela 2 de Pereira e Moreira (2021b), sem novos arredondamentos; cada ponto é uma classe de um participante. Cores: as dos conjuntos de estímulos (Figura 4.5); o cinza marca as comparações entre B e C. Os tracejados verdes mostram o que os autores esperavam.

O livro: dois experimentos online

O livro de Pereira e Moreira (2021a) refaz o estudo em outro formato. O Experimento 1 partiu da pergunta de se é possível transferir função de estímulos políticos para estímulos abstratos com uma diminuição progressiva do feedback, proposta em trabalho anterior de outros autores: todas as relações são treinadas juntas e o feedback diminui aos poucos, para reforçar de modo in-

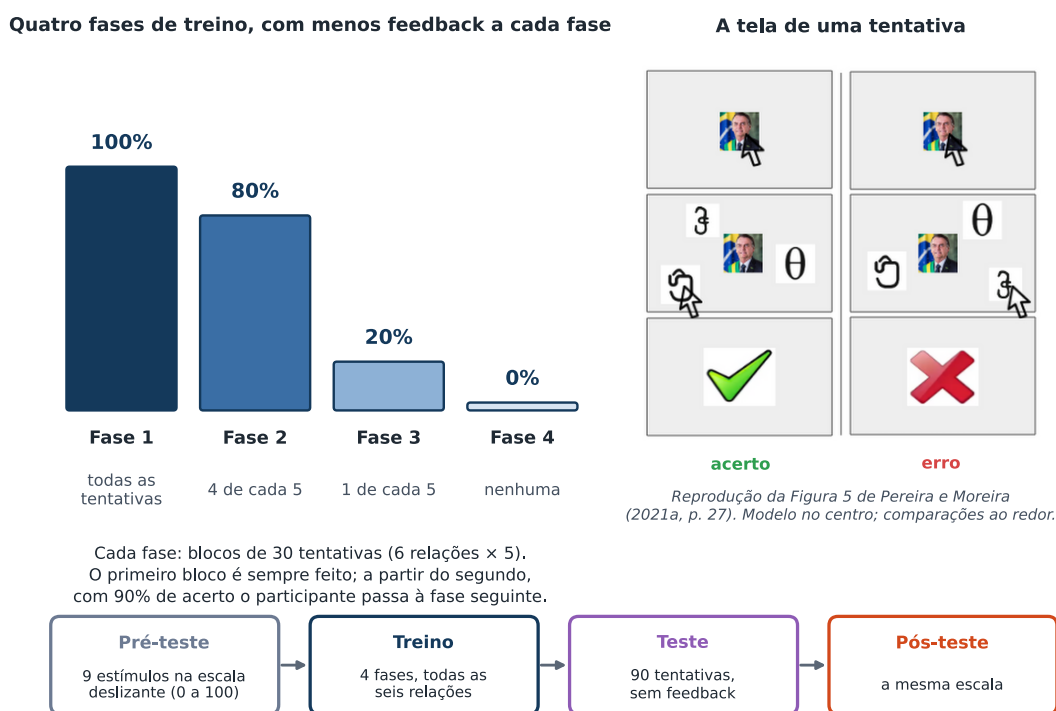
termitente e obter maior resistência à extinção no teste; os autores dizem haver apenas um estudo anterior com estímulos políticos nesse paradigma. Participaram cinco estudantes universitários de 18 a 25 anos, de cursos como Psicologia, Publicidade, Letras, Odontologia e Medicina, convidados por mensagem de WhatsApp em grupos de estudantes e escolhidos por conveniência; a coleta só começou depois da aprovação do comitê de ética e do termo de consentimento.

O experimento foi inteiramente online. Cada participante estava em casa, sozinho no cômodo, num computador com internet de pelo menos 10 Mbps; quem tivesse menos seria excluído. Uma plataforma apresentava as tentativas e os feedbacks de modo automático e aleatório e guardava os resultados. As instruções vinham em vídeos padronizados, com uma assistente virtual animada, em texto e em imagens; durante cada atividade, o pesquisador desligava áudio e câmera, e, entre as etapas, o participante podia fazer um intervalo de até cinco minutos. A escala tinha 13 pares de adjetivos bipolares e era respondida numa barra deslizante, de 0, mais perto do adjetivo negativo, a 100, mais perto do positivo, em vez dos sete pontos usuais. No pré-teste, cada participante avaliou nove estímulos, em ordem aleatória: três fotografias de presidentes (A1, Bolsonaro; A2, Lula; A3, Fernando Henrique Cardoso) e seis figuras abstratas (dois conjuntos de três); o botão para avançar só era liberado depois de 15 segundos e de avaliados todos os pares.

O treino tinha quatro fases, com o feedback cada vez menor (Figura 4.10). Em cada tentativa, o modelo aparecia no centro da tela, e, depois do clique nele, três comparações apareciam ao redor; o acerto era seguido de um ícone verde com um “v” e o erro, de um ícone vermelho com um “x”, e depois de um segundo de tela branca. Na primeira fase, o feedback vinha em todas as tentativas (100%); na segunda, em quatro de cada cinco (80%); na terceira, em uma de cada cinco (20%); na quarta, em nenhuma (0%). Em cada fase, cada uma das seis relações era apresentada cinco vezes, em blocos de 30 tentativas em ordem aleatória. O primeiro bloco era sempre feito; a partir do segundo, com ao menos 90% de acerto o participante passava à fase seguinte, e, se não atingisse o critério em 30 minutos, o sistema encerrava a tarefa e o participante era desqualificado. Por fim, o teste de simetria e transitividade era um único bloco de 90 tentativas randomizadas, sem feedback, com cinco tentativas por relação: as seis da linha de base, as de simetria (B e C como modelo, A como comparação) e as de transitividade (BC e CB).

FIGURA 4.10

O experimento online do livro: quatro fases de treino e a tela de cada tentativa



Nota. A tela é reprodução da Figura 5 de Pereira e Moreira (2021a); o restante é esquema do autor, a partir do método da obra.

O que o livro encontrou

Com o feedback caindo, quase todos os erros do treino aconteceram quando ele era total. Somando as Tabelas 1 a 5, do Experimento 1, e 27 a 31, do Experimento 2, houve 115 erros, dos quais 104 na fase de 100% de feedback, cinco na de 80%, dois na de 20% e quatro na de 0% (a soma é do autor deste livro; Figura 4.11). O maior número de erros foi o de P5, com 36 erros nas 120 tentativas que fez na primeira fase, e o de P1, com 20 erros em 90 tentativas; nas outras fases, todos fizeram 60 tentativas e erraram no máximo duas.

No teste do Experimento 1, os erros foram raros: P1 teve um (uma relação de transitividade, c1b1), P2 nenhum, P3 dois (ambos em b1c1), P4 dois (b2c2 e c1b1) e P5 nenhum, de modo que todos acertaram mais de 90% das relações em cada etapa do teste. Para saber se a avaliação foi transferida, os autores calcularam a semelhança entre as avaliações de dois estímulos de uma classe como 100 menos o valor absoluto da diferença entre as duas notas; quanto mais perto de 100, mais parecidas as avaliações, e a diferença entre a semelhança do pós-teste e a do pré-teste, positiva se a semelhança aumentou, mos-

tra o que o treino fez. As relações são chamadas ab (fotografia e primeiro conjunto de figuras), ac (fotografia e segundo conjunto) e bc (entre os dois conjuntos de figuras).

FIGURA 4.11

Erros de cada participante em cada fase de treino

Erros em cada fase de treino (porcentagem de tentativas com feedback)					
	100%	80%	20%	0%	
P1	20	0	0	1	Experimento 1 fotografias e figuras abstratas
P2	13	1	0	0	
P3	2	0	1	1	
P4	6	1	0	0	
P5	36	0	0	0	
P6	6	2	1	0	Experimento 2 faces neutras e figuras abstratas
P7	2	0	0	1	
P8	3	0	0	0	
P9	5	1	0	0	
P10	11	0	0	1	
Total	104	5	2	4	

Na fase de 100%, P1 fez 90 tentativas, P5 fez 120 e os demais 60; nas outras fases, 60 cada.

Nota. Números das Tabelas 1 a 5 e 27 a 31 de Pereira e Moreira (2021a); os totais por fase são do autor deste livro.

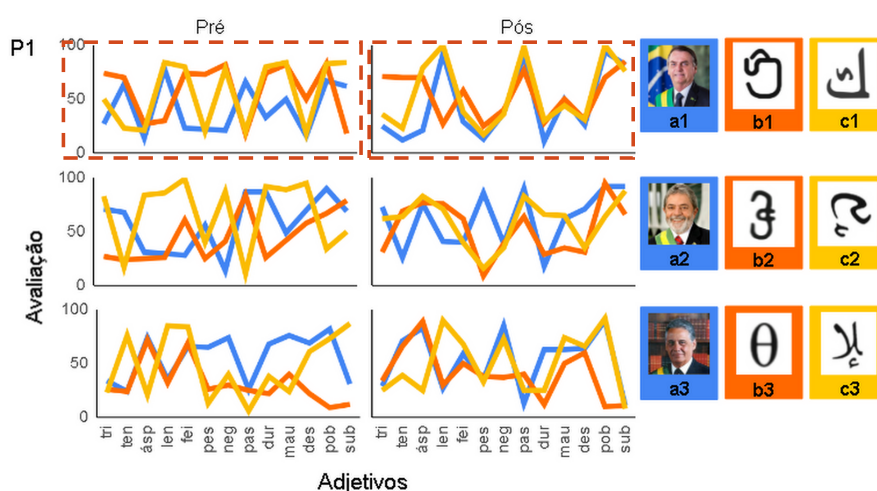
A Figura 4.12 reproduz a Figura 11 do livro, com as avaliações de P1 nas três classes, no pré e no pós-teste. As linhas azuis são as notas do estímulo a (a fotografia); as laranja, as do b; as amarelas, as do c, as mesmas cores dos conjuntos deste capítulo; os 13 pares de adjetivos estão no eixo horizontal, e as notas, de 0 a 100, no vertical. Os autores descrevem que, para P1, as avaliações dos estímulos abstratos ficaram mais parecidas com as do modelo na relação a1b1c1 no pós-teste do que no pré-teste; é o que se vê nos dois quadros tracejados, em que as três linhas se afastam no pré-teste e seguem mais juntas no pós-teste. Como os pares de adjetivos tinham polos alternados, uma limitação que os autores registram, o que se compara na figura é a distância entre as linhas, e não a altura de cada uma.

Na Tabela 26, o aumento médio da semelhança foi de 9,4 pontos percentuais na relação ab, 14,1 na ac e 6,0 na bc (Figura 4.13). Os valores por participante variam: as médias da relação ac ficaram entre 8,8 e 20,9, as da ab entre 4,8 e 12,8, e a da bc foi negativa em um participante, P5 (-0,9), isto é, a semelhança

entre as duas séries de figuras diminuiu um pouco para ele. Os autores observam também que, com exceção de P3, as avaliações das fotografias mudaram menos do pré para o pós-teste do que as das figuras abstratas, como esperavam, por os presidentes já terem um conceito prévio; para P3, a semelhança de duas relações já existia no pré-teste, e por isso dizem que, apenas pelos gráficos, é difícil afirmar a transferência de função.

FIGURA 4.12

Avaliações de P1 no pré e no pós-teste, nas três classes (Experimento 1 do livro)



- 1 Cada linha colorida é um estímulo: azul, a fotografia (a); laranja, a figura b; amarela, a figura c.
- 2 Cada ponto é um par de adjetivos (13 pares); a altura é a nota, de 0 a 100.
- 3 Compare os dois quadros tracejados (classe 1, Bolsonaro): depois do treino, as três linhas ficam mais próximas (livro, p. 48).

Reprodução da Figura 11 de Pereira e Moreira (2021a, p. 48).

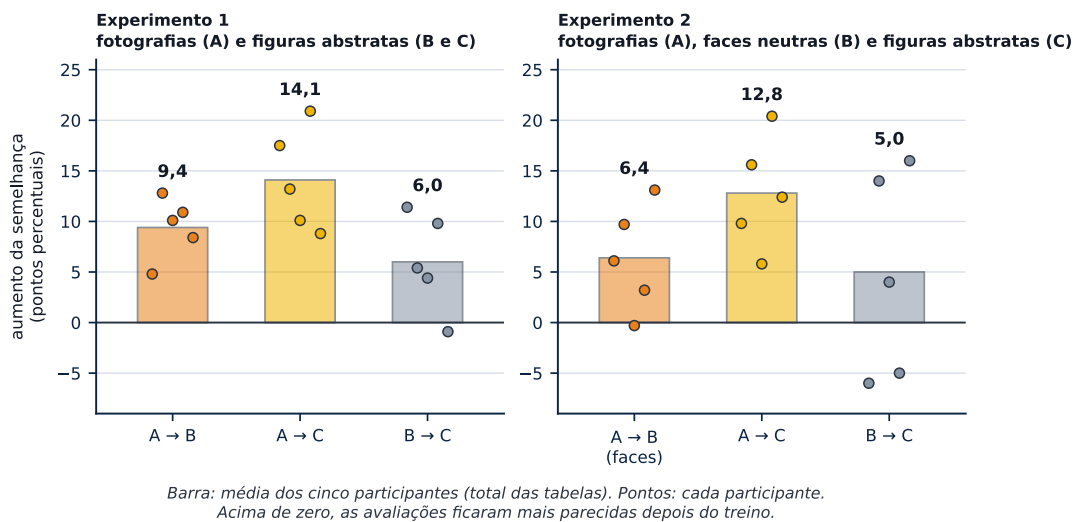
Nota. Reproduzida de Pereira e Moreira (2021a), com o quadro tracejado e o guia de leitura acrescentados.

O Experimento 2 parte de uma hipótese: quanto maior a carga afetiva pré-experimental dos estímulos, maior a dificuldade de formar classes e de transferir função. Os autores a ilustram com um estudo sobre ciúme, em que sete de 17 universitários não formaram classes, e se propõem verificar a formação de classes e a transferência com faces de pessoas comuns, avaliadas como neutras, e investigar se a transferência é mais fácil para estímulos abstratos do que para figuras humanas. Os cinco participantes tinham perfil e recrutamento idênticos aos do primeiro experimento. O conjunto B foi composto por faces de pessoas desconhecidas, de raça branca, escolhidas por terem sido as avaliadas como mais próximas de neutras em Cameoka e Moreira (2021), numa escala de 10 pontos (B1 e B2 com média 5,5 e B3 com 5,4); o conjunto C eram figuras abstratas, e o procedimento foi idêntico ao do Experimento 1.

No teste do Experimento 2, P6 teve seis erros em 90 tentativas (um de identidade, um de simetria e quatro de transitividade, 93,33% de acerto geral), P7 e P10 nenhum, P8 um e P9 um. Na Tabela 52, o aumento médio da semelhança foi de 6,4 pontos percentuais na relação ab, entre as fotografias e as faces; de 12,8 na ac, entre as fotografias e as figuras abstratas; e de 5,0 na bc. O aumento na ac foi o dobro do da ab. Por participante, P8 teve média de -0,3 na ab e 5,8 na ac, e, em seu gráfico, as avaliações de comparação ficaram neutras também no pós-teste; em P9, a transferência foi sobretudo da fotografia de Lula para as figuras da classe 2, menos nas classes 1 e 3, e não foi identificável para as faces.

FIGURA 4.13

Aumento da semelhança entre as avaliações, do pré para o pós-teste, nos dois experimentos



Nota. Valores das Tabelas 26 e 52 de Pereira e Moreira (2021a): barras, o total dos cinco participantes; pontos, a média de cada um. Cores: as dos conjuntos de estímulos, com o cinza para a relação entre B e C.

Para os autores, o aumento maior na relação ac do que na ab, de 50% no primeiro experimento e de 100% no segundo, indica que é mais fácil mudar a atitude para símbolos abstratos do que para faces neutras. Propõem que as faces, mesmo escolhidas por serem neutras, têm características, como ser de um homem, branco, de barba e cabelo, que lhes dão carga pré-experimental maior do que a dos estímulos abstratos, o que pode ter dificultado a transferência. As limitações que registram são a coleta online, dependente da velocidade da internet e feita em ambientes domésticos, e os pares de adjetivos com polos alternados; cada experimento foi feito numa sessão de uma hora a uma hora e meia.

O que os dois trabalhos permitem dizer

Dois achados do artigo importam para o argumento do capítulo. O primeiro é que as siglas dos partidos e as palavras *direita*, *esquerda* e *centro* foram relacionadas corretamente às figuras abstratas nos testes, embora nenhuma discriminação condicional com elas tivesse sido programada; para os autores, isso sugere que já faziam parte, antes do experimento, de uma classe com as fotografias dos presidentes. No pré-teste, os participantes avaliaram sistematicamente as fotografias de Bolsonaro, Lula e Fernando Henrique Cardoso como, respectivamente, direita, esquerda e centro. A leitura que o capítulo faz disso, a partir da inferência dos autores, é que o rótulo ideológico faz parte, para esses participantes, da classe com que o nome do político já chega ao experimento. O segundo achado é um cuidado dos autores: avaliações neutras, em torno do centro da escala, não são ausência de significado ou de atitude, e P02 deu avaliações positivas e negativas às figuras abstratas no pré-teste, o que sugere significado pré-experimental para elas.

Os autores também apontam o que falta. Propõem replicar o procedimento com outras estruturas de treino, séries lineares em vez de um-para-muitos, para avaliar o efeito da distância nodal, já que a semelhança esperada entre B e C vale em parte porque a distância de A a B e de A a C era a mesma. E sugerem replicar o trabalho com fotografias de pessoas anônimas no lugar das figuras abstratas, o que demonstraria a formação de atitudes em relação a pessoas com quem o participante nunca interagiu, fenômeno diretamente ligado a preconceitos, estereótipos e discursos de ódio. Esse é, em linhas gerais, o desenho do Experimento 2 do livro, com faces em lugar de figuras.

■ DADO

Em cinco estudantes, depois de aprenderem relações entre fotografias de três presidentes e figuras abstratas, as avaliações das figuras abstratas no diferencial semântico ficaram mais próximas das da fotografia correspondente; o Grau de Equivalência entre fotografia e figura B passou de 90% nos cinco participantes e nas três classes (Pereira & Moreira, 2021b). No livro dos mesmos autores, em dez participantes em dois experimentos, o aumento da semelhança foi menor para faces neutras do que para figuras abstratas (Pereira & Moreira, 2021a).

■ INTERPRETAÇÃO

A avaliação que cada participante já fazia de cada político passou a ser emitida também diante de estímulos sem história com ele, depois de uma sessão de cerca de uma hora e meia. O delineamento é de pré e pós-teste com sujeito único, sem condição de controle sem treino; os autores não sabem dizer se parte das diferenças em estímulos com significado prévio é variação de medidas repetidas.

■ HIPÓTESE

Que a transferência seja menor para estímulos de maior carga afetiva prévia, como sugerem os autores desse livro, é compatível com Haydu et al. (2015) e com Merwin e Wilson (2005), mas não foi manipulado: faces e figuras abstratas diferem em várias dimensões. Que o rótulo direita-esquerda seja parte da classe com que cada nome chega ao eleitor é indicação deste estudo, em cinco estudantes de Psicologia, não um dado sobre eleitores. O Capítulo 7 (seção 7.6) retoma a ideia para nomes e números de candidatos.

4.6 “Nós” e “eles”: relações de oposição e de comparação

A equivalência é uma relação de coordenação, mas não é a única que os humanos derivam. Dixon et al. (2003) lembram que padrões confiáveis de responder relacional, quando ficam sob controle contextual de pistas do ambiente (como as palavras “mais que”), são chamados de molduras relacionais, e citam igualdade, oposição, diferença, mais que e menos que, antes e depois. Weinstein et al. (2008) fazem o mesmo ponto: a equivalência não é a única relação funcional possível na transformação de funções. A diferença importa porque, numa relação que não é de equivalência, os estímulos não passam a compartilhar funções equivalentes; por isso os autores preferem “transformação” a “transferência”, que permite uma descrição mais ampla.

CONCEITO EM USO

Relações derivadas de oposição e de comparação (Dixon et al., 2003)

Uma moldura relacional é um padrão de responder relacional sob controle contextual de pistas do ambiente (por exemplo, “é o oposto de”, “é maior que”). Na *oposição*, dois estímulos são o oposto de um terceiro: se A é o oposto de Y e A é o oposto de Z, então Y é o mesmo que Z (Dixon et al., 2003; exemplo ilustrativo dos autores, sem dados). Na *comparação* (mais que, menos que), os estímulos relacionados não ficam com as mesmas funções: a relação define como a função se transforma, e não apenas se ela é compartilhada (Weinstein et al., 2008).

Os estudos de equivalência com estímulos sociais trazem pistas de que a oposição opera. Em McGlinchey e Keenan (1997), os participantes que não responderam por equivalência foram entrevistados: escolheram estímulos novos porque os que formariam a relação esperada eram vistos como opostos e, por isso, não combinavam. Alguns argumentaram que a dançarina irlandesa e a bandeira britânica não tinham nada em comum e que, portanto, a outra comparação “devia ser” a combinação. Os autores levantam que esses participantes podem ter respondido sob uma moldura de oposição, e relatam que um estudo anterior, de Moxon et al. (1993), tentou treinar classes com estímulos estereotipicamente vistos como opostos, e vários sujeitos não responderam por equivalência; a descrição confere com o artigo de Moxon et al.. O dado é modesto: algumas justificativas verbais, em entrevistas, depois do teste.

É um bom ponto para ler “nós” e “eles”. Em Haydu et al. (2015), as palavras foram distribuídas para conflitar com a história de cada torcida: o emblema do clube apoiado com *ruim* e o do principal rival com *bom*; os erros mais frequentes foram, de novo, escolher o que a história prescrevia. Que a história de uma torcida seja uma história de oposição entre “nós” e “eles” é coerente com o desenho dos autores, que escolheram essa população por causa de rivalidades declaradas entre as torcidas.

■ DADO	Em McGlinchey e Keenan (1997), participantes que não responderam por equivalência justificaram, em entrevista, que os estímulos da relação esperada eram vistos como opostos.
■ INTERPRETAÇÃO	Quando a classe ensinada põe lado a lado estímulos que a comunidade verbal trata como opostos, parte dos participantes responde conforme a oposição, e não conforme a coordenação pedida pela tarefa.
■ HIPÓTESE	No discurso político, a distinção “nós”-“eles” funcionaria como relação de oposição: atribuir um atributo aversivo a “eles” transformaria a função de “nós” no sentido inverso, e dois grupos opostos a um mesmo “eles” passariam a ser tratados como equivalentes entre si (regra do exemplo de Dixon et al., 2003). Não há dado, neste capítulo, sobre discurso; a hipótese precisa de falas datadas, como as da ficha do Capítulo 3, e de teste com o procedimento de oposição.

Duas observações de Dixon et al. (2003) ajudam a antecipar o Capítulo 12, e valem, como tudo neles, como argumento sem dados. Instruções do tipo “não acredite que todos os membros do grupo X são Y” podem aumentar a saliência do grupo como conceito categórico e acrescentar atributos à categoria. E, na análise dos autores, preconceito e atos de terror envolvem a transformação

derivada das funções de indivíduos ou grupos com base em contato frequentemente indireto com poucos membros do grupo. A seção 5.4 usa essa leitura para o rótulo de grupo.

4.7 Mudança de atitude: reorganização das classes e histórias em quadrinhos

Em termos do paradigma, mudar uma atitude é reorganizar classes. Cameoka e Moreira (2021) descrevem a reversão de classes de equivalência como um tipo de reorganização em que um rearranjo das contingências na linha de base produz mudanças no responder coerentes com o rearranjo. A Figura 4.14 esquematiza o procedimento com classes formadas no laboratório; nos estudos com faces, a classe a ser desfeita é a pré-experimental, e o treino estabelece uma classe conflitante (seção 4.4).

FIGURA 4.14

Mudança de atitude por reorganização de classes de equivalência

1. Antes: duas classes formadas



2. Treino de reversão

B1 → C2 e B2 → C1 passam a ser reforçadas; B1 → C1 e B2 → C2 deixam de ser.
"X" e "Y" não aparecem no treino: só as figuras B e C.

3. Depois, se houve reorganização (testado sem consequências)



Competição: a relação de "X" com o atributo negativo, mantida fora do laboratório pela comunidade verbal, continua disponível nos testes e pode impedir ou enfraquecer a reorganização.

Nota. Esquema original do autor, sem dados, no formato de reorganização de classes formadas em laboratório (como em Rosendo & de Melo, 2018). Nos estudos de Carvalho e de Rose (2014) e Mizael et al. (2016a), o treino AB e BC cria uma classe que conflita com a história pré-experimental.

A evidência é mista, e os resultados mais bem sustentados são os de reorganização parcial. Rosendo e de Melo (2018) estudaram classes com estímulos abstratos, adjetivos e profissões associadas aos gêneros masculino e feminino. Na Etapa 1, dois grupos experimentais de oito participantes cada foram expostos a treinos e testes de formação de classes por pareamento ao modelo com atraso, e a formação de classes se verificou nos dois. Na Etapa 2, só o Grupo Experimental 1 recebeu treinos de reversão, que rearranjam as contingências aprendidas na Etapa 1, e testes de reorganização de classes. Um grupo de controle, de 32 participantes, não passou por treino nem teste algum. Os três grupos avaliaram as profissões e os símbolos abstratos num diferencial semântico. No Grupo Experimental 1, os participantes atingiram o critério de reorganização, de 91% de acertos, apenas nas relações de simetria; nos demais testes, os escores ficaram próximos de 83% para seis participantes. Em comparação com o controle, as avaliações no diferencial semântico foram parcialmente coerentes com as classes e com os treinos de reversão, com transferência de função para algumas escalas, e as autoras concluem que é preciso melhorar os controles sobre a história pré-experimental com os estímulos e sobre a quantidade de relações revertidas.

Com faces e crianças, a tabela da seção 4.4 já trouxe a variação: uma de quatro crianças em de Carvalho e de Rose (2014), 13 de 13 com treino modificado em Mizael et al. (2016a), 33 de 46 em Mizael et al. (2021b). Dois pontos interessam à leitura política. Um é a manutenção: em Mizael et al. (2021b), 27 das 46 crianças mantiveram ao menos uma relação depois de seis semanas, e o grupo controle manteve menos. O outro é o tempo: Mizael et al. (2016b) sustentam que um treino de uma ou duas horas é, provavelmente, insuficiente para modificar relações pré-experimentais amplamente reforçadas pela comunidade verbal. É uma hipótese dos revisores, não dado de um estudo que a tenha testado.

Histórias em quadrinhos como procedimento de mudança

Os estudos do Instituto Walden⁴ trocam o treino de classes, quando o objetivo é mudar uma atitude, por uma história em quadrinhos. A razão, para os autores de um dos volumes (Moreira et al., 2022a), é a resistência: os métodos tradicionais de mudança de atitude desconsideram a resistência atitudinal, que os métodos baseados em narrativas consideram, e as narrativas, por tratar o preconceito de forma mais sutil, diluída no enredo, têm a capacidade de minimizar a percepção de intenção persuasiva e permitem que o leitor se identifique com os personagens. Os autores não encontraram, na literatura, experimentos sobre a formação de atitudes depois da leitura de quadrinhos com o paradigma de equivalência e o diferencial semântico, e é essa a lacuna que o Estudo 1 do volume se propõe a ocupar.

Esse estudo, de Elisa de Pennafort Rabelo, tem o desenho que os seguintes repetem. Sete estudantes de qualquer curso superior, exceto Psicologia, em que só do primeiro semestre, participaram de duas sessões, uma para cada história, com intervalo de no máximo três dias, pelo Google Meet e pelo Google Forms (Moreira et al., 2022a). Cada história, feita no Pixton, tem sete quadrinhos e três personagens mulheres: duas em conflito de convivência na casa e uma em papel neutro; na primeira história, Amanda é a única personagem negra. Antes e depois da leitura, o participante avaliou os três personagens de cada história (neutro, vítima e agressor) numa escala de diferencial semântico de 13 pares de adjetivos, de 1 a 7. A leitura produziu efeito nas avaliações dos personagens de todos os participantes, e a maioria atribuiu valores mais positivos às vítimas, mais neutros aos personagens neutros e mais negativos aos agressores, sobretudo depois da segunda história, o que pode indicar interferência da primeira sessão. Os autores não puderam verificar a influência de vieses raciais e propõem, para isso, reproduzir o experimento com novos participantes, trocando a ordem de apresentação das histórias. A Figura 4.17 mostra, na primeira coluna, o elenco desse estudo, e a Figura 4.18, o conflito da primeira história.

O Estudo 2 do mesmo volume, de Ingrid Taveira Premuli, usa quadrinhos num tema que interessa à leitura política: o autoritarismo. As histórias tinham personagens que representavam os três fatores da Escala de Autoritarismo de Direita (autoritarismo, contestação à autoridade e tradicionalismo), e o estudo investigou, por pareamento ao modelo, a transferência de função de estímulos com significados de autoritarismo para estímulos abstratos. Segundo o resumo, 100% da amostra formou classes de equivalência e 80% tiveram as avaliações influenciadas pelas histórias; a ambivalência na avaliação, em alguns participantes, leva os autores a questionar a construção e a validade da escala de diferencial semântico (Moreira et al., 2022a).

O estudo político da série é o Experimento 8 da Parte I (Moreira et al., 2022c), de Luiz Renato Valio Freire, e a Figura 4.15 o mostra por inteiro. Cinco adultos leram, em três trechos, uma história de 18 quadrinhos com três personagens: C1, com ideal político voltado para a esquerda; C2, para a direita; e C3, com opiniões neutras, que serve para observar variáveis que estejam interferindo na pesquisa. Os personagens e os quadrinhos foram feitos no Storyboard Maker, e a história usa poucos cenários, com poucos elementos, para reduzir influências alheias ao que se estuda. No primeiro trecho (Q1), as três personagens se encontram e conversam sobre o tempo e a possibilidade de chuva; é a ambientação, e a avaliação que se segue mede a primeira impressão. No segundo (Q2), C1 e C2 expressam opiniões opostas sobre o armamento civil e vão a passeatas opostas, uma a favor e outra contra. No terceiro (Q3), as três divi-

dem um carro de aplicativo e duas bandeiras de movimentos político-ideológicos aparecem em cena, que, na figura, se leem como a bandeira do Brasil junto de C2 e a da foice e do martelo junto de C1. Depois de cada trecho, o participante avalia C1, C2 e C3 nos mesmos 13 pares de adjetivos, de 1 a 7. A comparação entre os três testes é a medida da mudança, com o sujeito como o próprio controle. Cada participante levou em média 20 minutos, numa coleta de uma semana, e o experimento começou só depois da aprovação do Conselho de Ética.

O método tem uma ambiguidade, que convém registrar. A primeira etapa é chamada de pré-teste e descrita como a apresentação das faces dos personagens seguida da escala, antes de o participante saber da ideologia política deles, mas a descrição dos trechos diz que Q1 foi usado antes da primeira avaliação. A Figura 4.15 segue esta segunda descrição, a mais detalhada; nas duas, o primeiro teste vem antes de qualquer elemento político, que a descrição dos estímulos exclui da primeira avaliação.

Para os gráficos, os autores converteram as marcações de 1 a 7 em valores de -3 a +3 e reorganizaram os pares que tinham sido apresentados invertidos, de modo que os adjetivos negativos ficassem na parte inferior; a linha azul é o Teste 1, a vermelha, o Teste 2, e a amarela, o Teste 3. A Figura 4.16 reproduz o gráfico do Participante 1. Em C2, as linhas dos Testes 2 e 3 quase se sobrepõem: o que mudou depois das opiniões políticas se manteve depois das bandeiras; na maior parte dos pares, as duas ficam abaixo de zero, e em agradável-desagradável chegam a -3. Em C1 e C3, as mudanças foram poucas, com exceções, e, no conjunto, o participante tendeu a avaliar C1 de forma mais positiva e C2 e C3 de forma mais negativa ao longo das três avaliações.

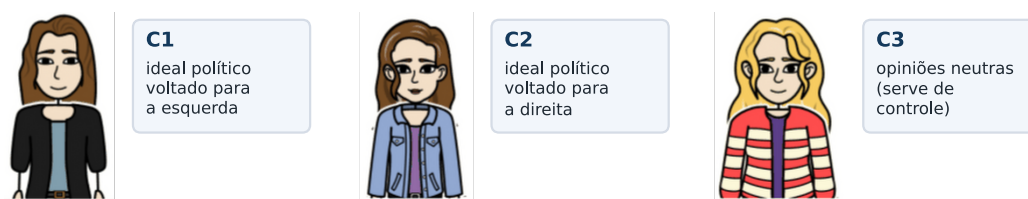
Entre os cinco participantes, houve mudanças de marcação para todas as personagens. C1 recebeu avaliações mais positivas que as outras duas, e C3, a personagem neutra, recebeu avaliações majoritariamente negativas dos Participantes 2, 3, 4 e 5. Os Participantes 3 e 5 mudaram de um polo da escala a outro, e o 2 manteve avaliações mais parecidas entre os testes. Em conversas espontâneas depois da aplicação, que os autores trazem como evidências anedóticas, os participantes disseram que C3 foi avaliada de modo mais negativo por sua passividade e pela falta de interação com as outras personagens, e que a personagem com ideologia mais parecida com a do participante era melhor avaliada. Para a análise do comportamento, o primeiro relato é instrutivo: a personagem planejada como controle neutro não funcionou como controle, porque, para quem avalia, não agir também é algo que se avalia. O segundo sugere que a semelhança entre quem avalia e quem é avaliado pode contar tanto quanto o que a história mostra, e um estudo com cinco participantes e sem grupo de comparação não separa uma coisa da outra. Os autores veem o

resultado em acordo com os dois estudos anteriores do grupo, o de Rabelo e o de Premuli, de que a leitura de quadrinhos pode gerar mudanças de atitude em relação a personagens fictícios, e recomendam um método misto, com entrevistas individuais sobre os motivos de cada avaliação.

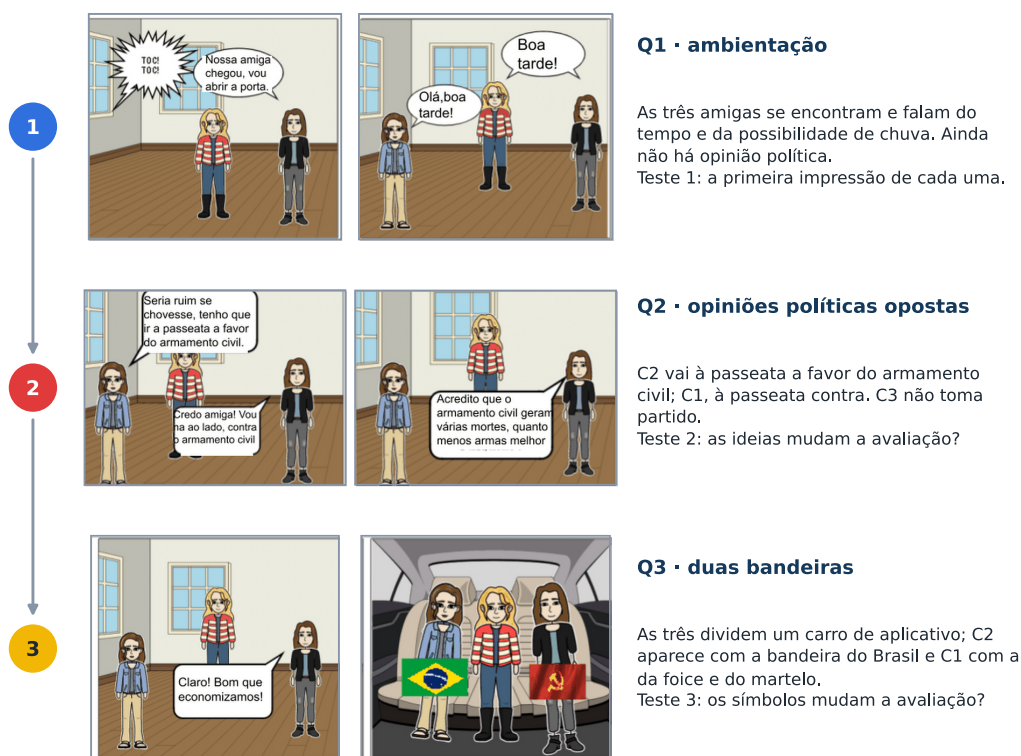
FIGURA 4.15

O Experimento 8 de Moreira et al. (2022c): três personagens, três trechos da história, três avaliações

1. O elenco: três personagens



2. Três trechos da história, três avaliações (13 pares de adjetivos, de 1 a 7)

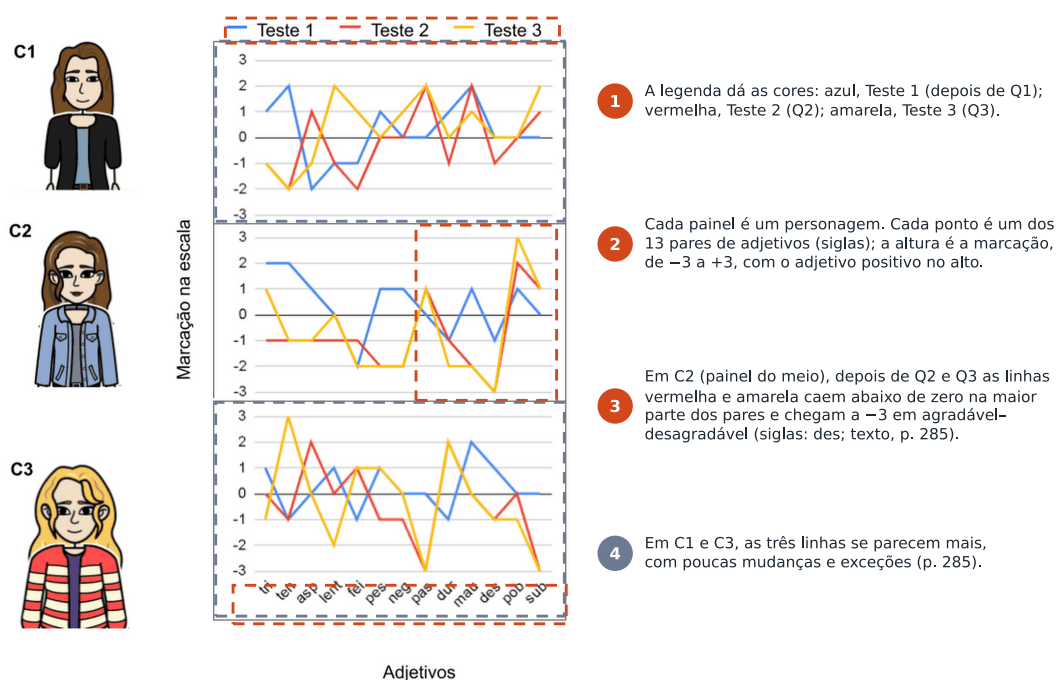


Em cada teste, o participante avalia C1, C2 e C3 nos mesmos 13 pares de adjetivos. A diferença entre os testes é a medida da mudança; cada participante é o seu próprio controle.

Nota. Personagens e quadrinhos reproduzidos de Moreira et al. (2022c); a organização em três momentos e os textos laterais são esquema do autor, a partir do método. As cores dos testes são as das linhas dos gráficos do estudo (Figura 4.16). A atribuição das bandeiras às personagens foi lida na última imagem; o texto do estudo diz só que são duas bandeiras.

FIGURA 4.16

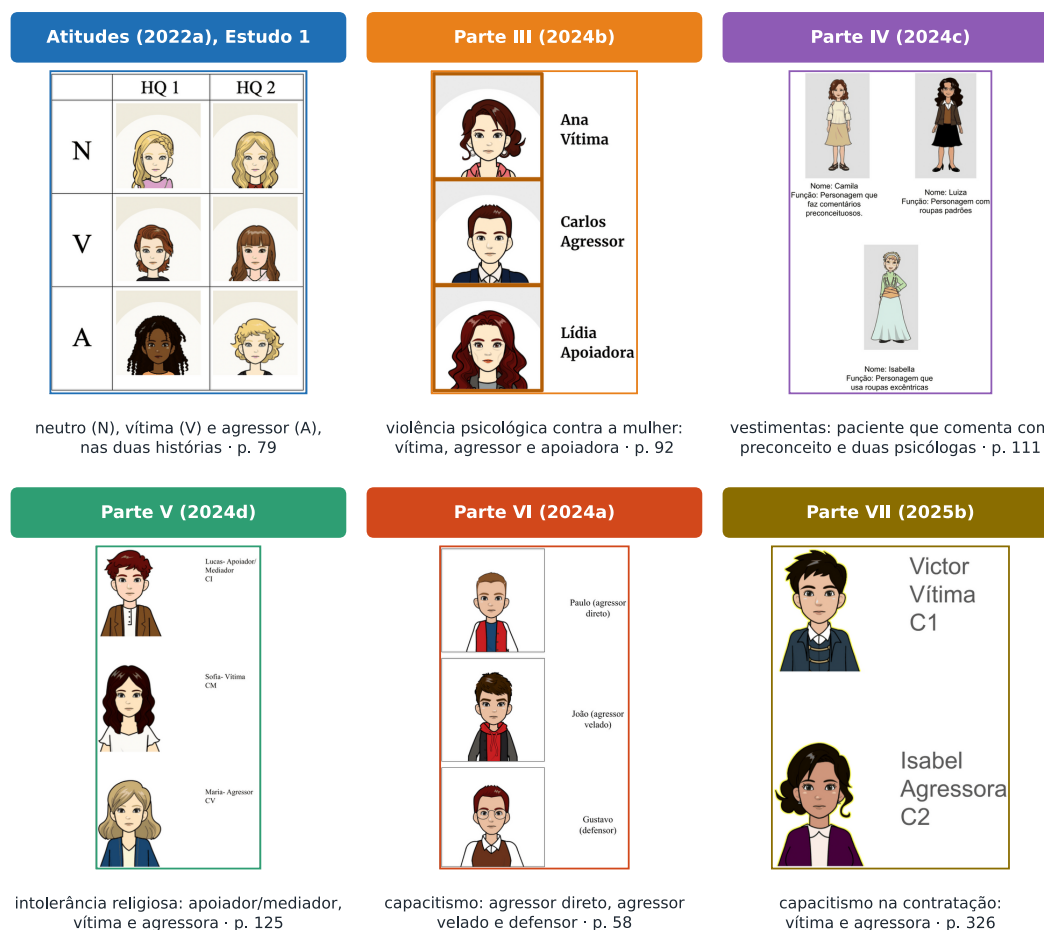
As avaliações do Participante 1 no Experimento 8



Reprodução da Figura 4 de Moreira et al. (2022c, p. 286).

Nota. Reprodução da Figura 4 de Moreira et al. (2022c), com os destaques e o guia acrescentados. As siglas do eixo são as do estudo (tri, ten, asp, lent, fei, pes, neg, pas, dur, mau, des, pob, sub), na ordem dos 13 pares de adjetivos, com o polo negativo nomeado; o estudo não as desdobra, e o texto lê “des” como agradável-desagradável.

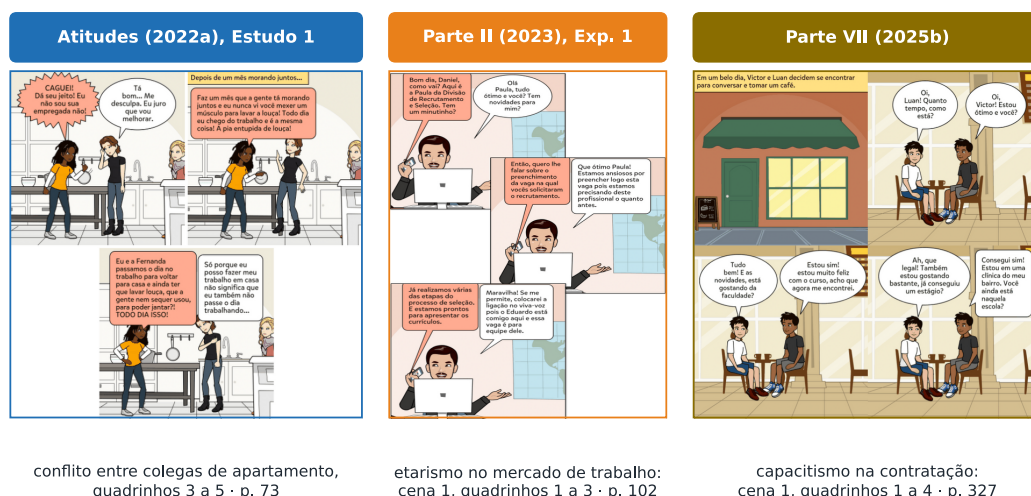
Os volumes seguintes repetem o desenho com outros temas e outros elencos. Nas Partes IV, V e VI, cada estudo traz, como o Experimento 8, personagens, enredo, roteiro, o formulário de diferencial semântico e testes antes e depois da leitura (Moreira et al., 2024c, 2024d, 2024a, sumários). A Tabela 4.2 lista os temas, e a Figura 4.17 mostra o elenco de seis estudos: pelos rótulos impressos nas figuras, ele costuma incluir quem sofre o ato, quem o pratica e quem apoia, defende, medeia ou observa. A Figura 4.18 mostra o que o participante lê: três cenas, de três livros.

FIGURA 4.17*O elenco de seis estudos com quadrinhos*

Nota. Personagens reproduzidos, com os rótulos de cada livro, de Moreira et al. (2022a; 2024b; 2024c; 2024d; 2024a; 2025b); as faixas e as legendas são do autor. Os temas são os dos títulos dos estudos (Tabela 4.2).

FIGURA 4.18

Três cenas de histórias em quadrinhos



Nota. Reproduzidas de Moreira et al. (2022a), do Experimento 1 de Moreira et al. (2023) e de Moreira (2025b), nesta ordem. O que os personagens dizem importa menos aqui do que a forma: poucos cenários, falas curtas e um conflito que o participante lê antes de avaliar os personagens.

TABELA 4.2

Livros do Instituto Walden4 com estudos de formação e mudança de atitudes e os temas dos estudos

Livro	Temas dos estudos
Atitudes: histórias em quadrinhos, análises semânticas em mídias sociais e o paradigma de relações de equivalência de estímulos (Moreira et al., 2022a)	Leitura de quadrinhos e personagens fictícios (Estudo 1); autoritarismo expresso por personagens (Estudo 2); trabalho doméstico (Estudo 3); comportamento verbal no Twitter, movimento Black Lives Matter (Estudo 4, Capítulo 6)
Preconceito: doze experimentos e um paradigma (Moreira et al., 2022b)	Controle de estímulos e mensuração de atitudes (Experimento 1); transferência de função e equivalência com estímulos de significado social (Experimentos 2 a 12, seção 4.4)
Estudos sobre histórias em quadrinhos e atitudes, Parte I (Moreira et al., 2022c)	Autismo na escola e no mercado de trabalho; transfobia; postura científica ou pseudocientífica; psicólogos com credo religioso; atratividade física; ex-detentos; preconceito político (Experimento 8); xenofobia; mulheres nos jogos eletrônicos; cirurgia plástica e influenciadoras digitais; ostentação de riqueza por celebridades
Parte II (Moreira et al., 2023)	Validade e precisão do instrumento de diferencial semântico; etarismo no mercado de trabalho; comunicação agressiva, assertiva e passiva; obesidade infantil; vestimenta; TEA no mercado de trabalho; preconceito e intolerância religiosa; discriminação racial
Parte III (Moreira et al., 2024b)	Violência psicológica contra a mulher em ambiente universitário; pseudociências; apelo emocional; gêneros musicais; transfobia

Livro	Temas dos estudos
Parte IV (Moreira et al., 2024c)	Transfobia; homofobia; discriminação por vestimentas; gênero no esporte de alta performance; gordofobia
Parte V (Moreira et al., 2024d)	Xenofobia; comportamento preconceituoso e intolerância religiosos
Parte VI (Moreira et al., 2024a)	Esquizofrenia no mercado de trabalho; capacitismo contra pessoas autistas; inclusão de crianças com TEA; ambivalência numa narrativa em quadrinhos
Parte VII (Moreira, 2025b)	Homofobia e ambivalência; racismo e homofobia no futebol; notícias de fofoca; pessoas em situação de rua; mães em entrevistas de emprego; tratamentos alternativos para o autismo; estímulos emocionais; contratação de pessoas com TEA; mídias sociais; natação inclusiva; o caso Gabriela Pugliesi; ex-adictos

Nota. Temas conforme os sumários dos volumes; “Parte II” a “Parte VII” abreviam o título Histórias em quadrinhos e o estudo das atitudes. Dos estudos da tabela, só dois foram lidos em detalhe neste capítulo, o Estudo 1 de Moreira et al. (2022a) e o Experimento 8 de Moreira et al. (2022c).

O volume organizado por Moreira (2025a) reúne estudos de sujeito único, majoritariamente com universitários, com pareamento ao modelo e com quadrinhos, sobre sexismo, homofobia, etarismo, transfobia, gordofobia e preconceito contra pessoas com síndrome de Down e com autismo, em que a atitude é medida, sobretudo, por escala de diferencial semântico. O resumo do volume relata que o pareamento formou classes e transferiu funções avaliativas, em especial para estímulos inicialmente neutros ou pouco carregados emocionalmente, e que houve resistência à mudança de preconceitos profundamente enraizados ou associados a forte carga emocional (resumo do volume).

■ DADO

Em estudos de reorganização com estímulos de significado social, a reorganização foi parcial (Rosendo & de Melo, 2018, só para simetria), variou com os parâmetros de treino (Mizael et al., 2016a, 2021b) e se manteve em parte depois de seis semanas (27 de 46). Nos dois estudos com quadrinhos lidos em detalhe, houve alterações nas marcações de todos os participantes (Moreira et al., 2022a, Estudo 1: sete; 2022c, Experimento 8: cinco), e a personagem planejada como neutra no Experimento 8 recebeu avaliações majoritariamente negativas de quatro dos cinco.

■ INTERPRETAÇÃO

Reorganizar uma classe pré-experimental é possível em condições de treino otimizadas, mas o efeito é parcial, medido numa escala logo depois do treino, e não diz o que a pessoa dirá em outro contexto. Os estudos com quadrinhos examinam o efeito de informação sobre a avaliação de personagens fictícios; com o sujeito como seu próprio controle (Moreira et al., 2022c), a mudança nas marcações não se distingue, sozinha, de efeito de repetição da medida, e o Estudo 1 de Moreira et al. (2022a) registra possível interferência da primeira sessão sobre a segunda.

■ **HIPÓTESE** A mudança de atitude no discurso político dependeria, além do treino, de contingências que continuem a reforçar as relações novas e deixem de reforçar as antigas; as relações antigas seguem reforçadas pela comunidade verbal fora do laboratório. É o que o Capítulo 12 trata como intervenção, não como dado deste.

4.8 Medidas, interpretações rivais e limites

A medida de atitude mais frequente nos estudos deste capítulo é uma escala: de diferencial semântico, com pares de adjetivos bipolares, em sete pontos no artigo com presidentes e em barra deslizante de 0 a 100 no livro, e em sete pontos nos estudos com quadrinhos e com gênero e profissões; Likert, em Dixon e Lemke (2007) e em Cameoka e Moreira (2021). Os autores do estudo com presidentes a tomam como medida de atitude porque ela relaciona adjetivos a estímulos (Pereira & Moreira, 2021b). Há medidas de outro tipo: o teste de associação implícita, usado por Weinstein et al. (2008), em que o autorrelato não se correlacionou com a resposta, e medidas baseadas em latência, que Mizael et al. (2021a) incluem entre as contribuições brasileiras para a avaliação de vieses raciais. Nenhuma delas deixa de ser comportamento sob controle do procedimento que a produz.

Do ponto de vista analítico-comportamental, uma marcação na escala é comportamento verbal dirigido a um pesquisador. Guerin (1994) observa que escalas e perguntas dependem de quem pergunta e da cultura verbal em que a pesquisa é feita, porque atitudes dependem sempre de consequências sociais, e que a audiência funciona como estímulo discriminativo para relatar atitudes específicas, de modo que uma pergunta pode “mandar” a resposta. Daí a expectativa de inconsistências entre o que se declara e o que se faz, sobretudo no comportamento verbal. A consequência é direta para os experimentos acima: uma mudança na marcação é uma mudança no que a pessoa marca, diante daquele pesquisador e daquela escala, e não uma mudança demonstrada no que ela fará ou dirá em outras audiências.

O instrumento dos dois trabalhos com presidentes tem duas peculiaridades que pedem cautela. A primeira é que os números de um e de outro não se comparam diretamente. No artigo, a escala tinha sete lacunas por par, marcadas em papel e convertidas em números de -3 a +3; no livro, era uma barra deslizante de 0 a 100, respondida no computador, com pares de polos alternados, e a semelhança entre estímulos foi calculada de outro modo, como 100 menos a diferença entre as notas. O GE e o GT do artigo e a semelhança do livro medem, em linhas gerais, o quanto duas avaliações se parecem, mas as

contas são diferentes. A segunda é uma inconsistência do artigo: o método diz que, no par Direita–Esquerda, Direita ficou com -3 e Esquerda com $+3$, e a discussão descreve as fotografias de Bolsonaro e de Lula como, respectivamente, direita (valores próximos de 3) e esquerda (valores próximos de -3). Os sinais da discussão são os inversos dos do método. Este livro registra a divergência e não a resolve, porque não tem acesso aos dados brutos, e o ponto que o capítulo usa, o de que os participantes avaliaram sistematicamente as fotografias como direita, esquerda e centro, não depende do sinal.

A seleção de estímulos também controla a medida. Cameoka e Moreira (2021) fizeram três estudos, com 75, 40 e 30 universitários, em que avaliaram, numa escala de 1 (mau) a 10 (bom), fotografias de homens brancos e de homens negros. A diferença entre as médias dos dois grupos de faces foi significativa nos Estudos 1 e 2 e não no Estudo 3, em que se usaram as faces de brancos mais bem avaliadas e as de negros pior avaliadas nos estudos anteriores; e a média das faces de brancos e de negros aumentou, de modo significativo, do Estudo 1 para o 2, em que as faces estavam discretamente sorridentes. Os autores supõem demonstrar, com isso, o potencial de distorção de uma seleção de estímulos pouco rigorosa. A consequência é que “a atitude em relação a X” medida com uma fotografia de X é, em parte, a atitude em relação àquela fotografia.

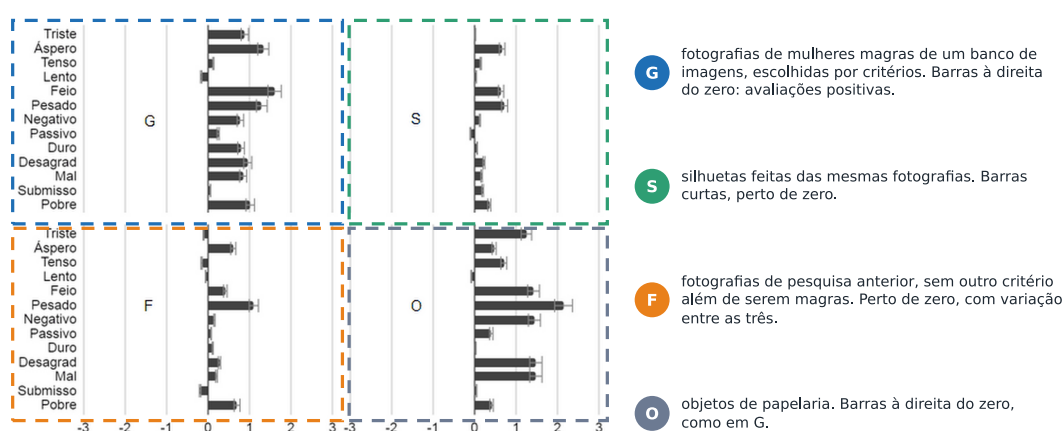
O Experimento 1A de Moreira et al. (2022b) mostra o problema num desenho feito para isso. Trinta e três pessoas, de 15 a 53 anos, avaliaram num formulário online 12 estímulos em 13 pares de adjetivos, de 1 a 7: três fotografias de mulheres magras de um banco de imagens, escolhidas por critérios (traje de lingerie ou biquíni, de frente, idade aparente de 20 a 30 anos, fundo branco e caráter comercial, e semelhança, julgada por oito juízes, com as silhuetas 2 ou 3 da Escala de Stunkard); três fotografias de mulheres magras de uma pesquisa anterior, sem outro critério; três silhuetas, feitas a partir das fotografias do primeiro conjunto, com o corpo pintado de preto; e três objetos de papelaria. Cada estímulo apareceu duas vezes, em ordem semialeatória. As fotografias escolhidas por critérios receberam avaliações positivas, iguais ou maiores que 0,8 na maior parte dos pares; as silhuetas, avaliações perto de zero, em geral entre 0 e 0,4; e a diferença foi estatisticamente significativa em todos os pares, exceto Tenso/Relaxado e Submisso/Dominante (Figura 4.19). Entre as três fotografias sem critério, uma recebeu avaliações mais positivas, outra mais neutras e a terceira mais negativas, e os autores atribuem a variação ao único critério usado, o de serem magras, a juízo do pesquisador. Como as silhuetas guardam o formato do corpo e ficaram perto de zero, os autores sugerem que as respostas às fotografias não estavam sob controle do formato do corpo, mas de outras partes dos estímulos, e concluem que a padronização na seleção dos estímulos é essencial para que as respostas fiquem sob controle da variável

programada, e que a cor da pele ou do cabelo, quando não é a variável de interesse, pode criar vieses avaliativos. Para pesquisas futuras, propõem dois estudos separados, um de seleção dos estímulos e outro para a tarefa de discriminação ou de reversão de equivalência. Para a leitura de falas políticas, a consequência é a mesma: o exemplar escolhido para representar um grupo é o que se avalia, e a atitude “em relação ao grupo” é uma inferência feita a partir dele.

FIGURA 4.19

O mesmo conceito, quatro tipos de estímulo, quatro perfis de avaliação

Média das avaliações de 33 participantes, em 13 pares de adjetivos (–3 a +3)



Reprodução da Figura 3 de Moreira et al. (2022b, p. 47).

Cada barra é a média de um par de adjetivos; à direita do zero está o adjetivo positivo. O que muda é o estímulo, não o conceito: G, S e F pretendiam representar mulheres magras (p. 41, 51).

Nota. Reprodução da Figura 3 de Moreira et al. (2022b), com os destaques e o guia acrescentados; cada barra é a média dos participantes para um par de adjetivos, e os traços finos são os da figura original (o texto não diz que medida representam). Os rótulos dos pares são os do gráfico.

Resta o que nenhuma dessas medidas alcança. Os experimentos tratam de comportamento verbal individual diante de estímulos escolhidos pelo pesquisador, em sessões curtas e com amostras pequenas, de 4 a 50 participantes nos estudos da Tabela 4.1 e de cinco por experimento nos de presidentes, muitas vezes estudantes universitários ou crianças recrutadas em escolas. Passarelli et al. (2023) distinguem o racismo em três níveis, estrutural, institucional e individual, e apontam a escassez de trabalhos experimentais, em parte pela dificuldade de defini-lo em termos comportamentais; o que este capítulo examinou é o nível individual. A passagem do laboratório ao discurso público é por analogia, e aqui ela é assumida como tal: o paradigma oferece um modelo de como funções avaliativas chegam a nomes sem contato direto e de como a história social compete com o que se ensina, e é com esse modelo que os Capítulos 5 e 7 leem as falas.

Interpretações rivais para o mesmo dado de transferência

Pareamento direto. No desenho de Pereira e Moreira (2021b), as figuras B e C foram relacionadas diretamente à fotografia (treinos AB e AC); a relação derivada é BC. O condicionamento avaliativo, em que o pareamento direto entre um estímulo neutro e outro com valência muda a resposta avaliativa ao neutro (Mizael et al., 2016b), explica a semelhança entre A e B, e entre A e C, sem recorrer à classe. Os autores propõem testar séries lineares contra um-para-muitos, com a distância nodal; Bortoloti e de Rose (2009) já mostraram que figuras a um nó das faces foram julgadas semelhantes e a três nós, não.

Controle pela audiência. A marcação pode refletir o que o participante supõe que o pesquisador espera (Guerin, 1994).

Linha de base não neutra. Avaliações neutras não são ausência de significado (Pereira & Moreira, 2021b), e P02 avaliou as figuras abstratas, no pré-teste, de modo positivo e negativo; a diferença entre pré e pós-teste mistura, então, transferência com o que as figuras já eram para ele.

Variação de medida repetida. Os autores admitem não saber quanto da diferença em estímulos com significado prévio é variação entre aplicações da escala.

Síntese

Neste livro, atitude é uma avaliação de um objeto que é, ao mesmo tempo, comportamento verbal: mantida por consequências arranjadas pela comunidade verbal e capaz de funcionar como tato, intraverbal ou mando. O paradigma de equivalência de estímulos oferece o modelo de como uma função avaliativa chega a um estímulo sem contato direto: depois que uma classe se forma por discriminações condicionais, a função de um membro aparece nos demais, em testes sem consequências diferenciais, e varia com a distância nodal.

A literatura com estímulos de significado social mostra um padrão. Com estímulos arbitrários, funções avaliativas aparecem em tamanho pequeno depois de treinos curtos. Com estímulos que contrariam a história do participante, a formação da classe é interrompida em parte dos casos, e a interrupção acompanha a força da história, que os estudos comparam (entre idades, entre sexos), mas não manipulam; com treinos modificados, crianças formaram a classe pretendida e a mantiveram em parte. Os experimentos com presidentes do Brasil mostram o mecanismo em estudantes universitários, num desenho de pré e pós-teste sem controle, com cinco participantes por experimento. Nos cinco do artigo, as avaliações das figuras abstratas passaram a se parecer com as das fotografias (GE acima de 90%), com mudanças moderadas (GT de 12% a 49% nas figuras B); nos dez do livro, a semelhança aumentou em média

de 5 a 14 pontos percentuais, mais para figuras abstratas do que para faces neutras. Os dois trabalhos sugerem que o rótulo direita-esquerda já faz parte da classe com que cada nome chega ao participante. Nos estudos do Instituto Walden⁴ com quadrinhos, as marcações mudaram em todos os participantes (sete em um estudo, cinco em outro), com o sujeito como o próprio controle e sem grupo de comparação; no estudo político, a personagem planejada como neutra não funcionou como controle, e um experimento sobre seleção de estímulos mostrou que o tipo de fotografia, e não só o conceito, controla a avaliação.

Todos esses dados são marcações em escala ou escolhas numa tela, dirigidas a um pesquisador, e admitem pelo menos duas leituras rivais: pareamento direto entre fotografia e figura, sem classe, e controle pela audiência. As relações de oposição entre “nós” e “eles” são uma hipótese promissora para ler o discurso político, apoiada por poucas entrevistas e sem dado sobre falas. O Capítulo 5 usa o que este capítulo estabeleceu sobre transferência para o rótulo de grupo, e o Capítulo 7, sobre nomes de candidatos.

Referências

- Ajzen, I. (2001). Nature and operation of attitudes. *Annual Review of Psychology*, 52, 27–58. <https://doi.org/10.1146/annurev.psych.52.1.27>
- Albarracin, D., & Shavitt, S. (2018). Attitudes and attitude change. *Annual Review of Psychology*, 69, 299–327. <https://doi.org/10.1146/annurev-psych-122216-011911>
- Barnes, D., Lawlor, H., Smeets, P. M., & Roche, B. (1996). Stimulus equivalence and academic self-concept among mildly mentally handicapped and nonhandicapped children. *The Psychological Record*, 46(1), 87–107. <https://doi.org/10.1007/bf03395165>
- Bohner, G., & Dickel, N. (2011). Attitudes and attitude change. *Annual Review of Psychology*, 62, 391–417. <https://doi.org/10.1146/annurev.psych.121208.131609>
- Bortoloti, R., & de Rose, J. C. (2009). Assessment of the relatedness of equivalent stimuli through a semantic differential. *The Psychological Record*, 59(4), 563–590. <https://doi.org/10.1007/bf03395682>
- Cameoka, M. C., & Moreira, M. B. (2021). Preconceito racial: viés na mensuração de atitudes produzido por controle de estímulos. *Acta Comportamentalia*, 29(1), 93–112. <https://doi.org/10.32870/ac.v29i1.78782>
- de Carvalho, M. P., & de Rose, J. C. (2014). Understanding racial attitudes through the stimulus equivalence paradigm. *The Psychological Record*, 64(3), 527–536. <https://doi.org/10.1007/s40732-014-0049-4>
- de Rose, J. C., McIlvane, W. J., Dube, W. V., Galpin, V. C., & Stoddard, L. T. (1988). Emergent simple discrimination established by indirect relation to differential consequences. *Journal of the Experimental Analysis of Behavior*, 50(1), 1–20. <https://doi.org/10.1901/jeab.1988.50-1>

- Dixon, M. R., & Lemke, M. (2007). Reducing prejudice towards Middle Eastern persons as terrorists. *European Journal of Behavior Analysis*, 8(1), 5–12. <https://doi.org/10.1080/15021149.2007.11434269>
- Dixon, M. R., Dymond, S., Rehfeldt, R. A., Roche, B., & Zlomke, K. R. (2003). Terrorism and relational frame theory. *Behavior and Social Issues*, 12(2), 129–147. <https://doi.org/10.5210/bsi.v12i2.40>
- Dougher, M. J., Augustson, E., Markham, M. R., Greenway, D. E., & Wulfert, E. (1994). The transfer of respondent eliciting and extinction functions through stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior*, 62(3), 331–351. <https://doi.org/10.1901/jeab.1994.62-331>
- Dymond, S., & Rehfeldt, R. A. (2000). Understanding complex behavior: The transformation of stimulus functions. *The Behavior Analyst*, 23(2), 239–254. <https://doi.org/10.1007/bf03392013>
- Guerin, B. (1994). Attitudes and beliefs as verbal behavior. *The Behavior Analyst*, 17(1), 155–163. <https://doi.org/10.1007/bf03392661>
- Haydu, V. B., Camargo, J., & Bayer, H. (2015). Effects of preexperimental history on the formation of stimulus equivalence classes: A study with supporters of Brazilian soccer clubs. *Psychology & Neuroscience*, 8(3), 385–396. <https://doi.org/10.1037/h0101276>
- Hayes, S. C., Kohlenberg, B., & Hayes, L. J. (1991). The transfer of specific and general consequential functions through simple and conditional equivalence relations. *Journal of the Experimental Analysis of Behavior*, 56(1), 119–137. <https://doi.org/10.1901/jeab.1991.56-119>
- McGlinchey, A., & Keenan, M. (1997). Stimulus equivalence and social categorization in Northern Ireland. *Behavior and Social Issues*, 7(2), 113–128. <https://doi.org/10.5210/bsi.v7i2.310>
- Merwin, R. M., & Wilson, K. G. (2005). Preliminary findings on the effects of self-referring and evaluative stimuli on stimulus equivalence class formation. *The Psychological Record*, 55(4), 561–575. <https://doi.org/10.1007/bf03395527>
- Mizael, T. M., & de Rose, J. C. (2017). Análise do comportamento e preconceito racial: possibilidades de interpretação e desafios. *Acta Comportamentalia*, 25(3), 365–377. <https://doi.org/10.32870/ac.v25i3.61632>
- Mizael, T. M., Coelho, C. L., Rodrigues, W. C., & de Almeida, J. H. (2021a). Racial issues and behavior analysis: Experiences and contributions from Brazil. *Behavior and Social Issues*, 30(1), 495–513. <https://doi.org/10.1007/s42822-021-00071-1>
- Mizael, T. M., de Almeida, J. H., Roche, B., & de Rose, J. C. (2021b). Effectiveness of different training and testing parameters on the formation and maintenance of equivalence classes: Investigating prejudiced racial attitudes. *The Psychological Record*, 71(2), 265–277. <https://doi.org/10.1007/s40732-020-00435-w>
- Mizael, T. M., de Almeida, J. H., Silveira, C. C., & de Rose, J. C. (2016a). Changing racial bias by transfer of functions in equivalence classes. *The Psychological Record*, 66(3), 451–462. <https://doi.org/10.1007/s40732-016-0185-0>
- Mizael, T. M., Dos Santos, S. L., & De Rose, J. C. C. (2016b). Contribuições do paradigma de equivalência de estímulos para o estudo das atitudes. *Interação em Psicologia*, 20(2). <https://doi.org/10.5380/psi.v20i2.46278>
- Moreira, M. B. (Org.). (2025a). *Formação e mudança de atitudes: Pareamento ao modelo e história em quadrinhos* (1ª ed.). Instituto Walden4.

- Moreira, M. B. (Org.). (2025b). *Histórias em quadrinhos e o estudo das atitudes - Parte VII*. Instituto Walden4. https://www.walden4.com.br/livros/pdfs/moreira_2025_hq7.pdf
- Moreira, M. B., Ávila, A., Corrêa, B. F., Reis, F. G. dos, Oliveira, L. I. R. de, Pol, L. V., Silva, L. C. da, & Aguiar, S. H. A. (2023). *Histórias em quadrinhos e o estudo das atitudes - Parte II* (1ª ed.). Instituto Walden4. https://www.walden4.com.br/livros/pdfs/moreira_etal_2023_hq2.pdf
- Moreira, M. B., Cardoso, G. P., Silveira, L. A., Carvalho, L. R., & Rabelo, M. M. (2024a). *Histórias em quadrinhos e o estudo das atitudes - Parte VI*. Instituto Walden4. https://www.walden4.com.br/livros/pdfs/moreira_etal_hq6_2024.pdf
- Moreira, M. B., Lemos, I. K., Nogueira, L., Silverio, J. P. N. A., Bacelar, J. M. C., Moraes, M., Godoy, M. B. C., & da Silva, V. A. L. R. (2024b). *Histórias em quadrinhos e o estudo das atitudes - Parte III*. Instituto Walden4. https://www.walden4.com.br/livros/pdfs/moreira_etal_hq3_2024.pdf
- Moreira, M. B., Martins, A. C. M., Silva, A. F. V., Ribeiro, G. M., Faria, G. M. S., & Bezerra, J. G. N. (2024c). *Histórias em quadrinhos e o estudo das atitudes - Parte IV*. Instituto Walden4. https://www.walden4.com.br/livros/pdfs/moreira_etal_hq4_2024.pdf
- Moreira, M. B., Pova, A. M. R., Belarmino, A. J. A., Meireles, A. D. R., & Tanabe, G. Y. R. (2024d). *Histórias em quadrinhos e o estudo das atitudes - Parte V*. Instituto Walden4. https://www.walden4.com.br/livros/pdfs/moreira_etal_hq5_2024.pdf
- Moreira, M. B., Premuli, I. T., Rabelo, E. de P., Pessoa, M. B. F., & Medeiros, M. R. G. C. (2022a). *Atitudes: Histórias em quadrinhos, análises semânticas em mídias sociais e o paradigma de relações de equivalência de estímulos* (1ª ed.). Instituto Walden4. https://www.walden4.com.br/livros/pdfs/moreira_etal_2022_q0.pdf
- Moreira, M. B., Rocha, G., Lago, J., Soares, A., Santos, B. dos, Simões, C., Dias, G., Nunes, J., Alves, K., Pires, M., Jardim, M., Chadud, L., Alvarenga, O., & Souza, W. (2022b). *Preconceito: Doze experimentos e um paradigma* (1ª ed.). Instituto Walden4. https://www.walden4.com.br/livros/pdfs/moreira_etal_2022_12exp.pdf
- Moreira, M. B., Sousa, C. P., Tomasini, C., Craveiro, H. do A., Rodrigues, I. L., Dutra, I., Barros, J. F., Freire, L. R. V., D'arcanchy e Vilela, M. E., Cautela, M. E. T. C., Oliveira, P. M. D. N., Soares, S. K. S., & Andrade, V. M. (2022c). *Estudos sobre histórias em quadrinhos e atitudes - Parte I* (1ª ed.). Instituto Walden4. https://www.walden4.com.br/livros/pdfs/moreira_etal_2022_hq1.pdf
- Moxon, P. D., Keenan, M., & Hine, L. (1993). Gender-role stereotyping and stimulus equivalence. *The Psychological Record*, 43, 381–394.
- Passarelli, D. A., Rico, A. S., & Silvestre, M. H. (2023). Contribuições da análise do comportamento para a compreensão do racismo. *Revista Brasileira de Análise do Comportamento*, 19(1). <https://doi.org/10.18542/rebac.v19i1.14941>
- Pereira, P. B. G., & Moreira, M. B. (2021a). *Atitudes em relação a presidentes do Brasil: Formação de classes de equivalência e transferência de função* (1ª ed.). Instituto Walden4. https://www.walden4.com.br/livros/pdfs/pereira_moreira_2021_presidentes.pdf
- Pereira, P. B. G., & Moreira, M. B. (2021b). Transferência de atitudes em relação a presidentes da república. *Acta Comportamentalia*, 29(4), 137–154. <https://doi.org/10.32870/ac.v29i4.80318>
- Rosendo, A. P., & de Melo, R. M. (2018). Transferência de função e reorganização de classes de equivalência relacionadas a gênero e profissões. *Revista Brasileira de Análise do Comportamento*, 14(1), 31–43. <https://doi.org/10.18542/rebac.v14i1.7157>

- Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. Matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, 37(1), 5–22. <https://doi.org/10.1901/jeab.1982.37-5>
- Weinstein, J. H., Wilson, K. G., Drake, C. E., & Kellum, K. K. (2008). A relational frame theory contribution to social categorization. *Behavior and Social Issues*, 17(1), 40–65. <https://doi.org/10.5210/bsi.v17i1.406>



PRÁTICAS COM EVIDÊNCIA

O que funciona no autismo, prática por prática

Uma página para cada prática com evidência, a partir da revisão do National Clearinghouse on Autism Evidence and Practice: 31.779 artigos triados e 972 analisados, com o número de estudos que sustenta cada uma.

28 práticas descritas

Acesse, de graça e sem cadastro

walden4.com.br/pratica-baseada-em-evidencias/autismo



5

Discurso de ódio: mando disfarçado e controle aversivo

Em 2011, o programa CQC, da TV Bandeirantes, perguntou ao então deputado Jair Bolsonaro o que ele faria se tivesse um filho gay. Segundo a Agência Lupa, que reproduz em seu relatório de 2026 a checagem publicada em 3 de agosto de 2018, ele respondeu: “Isso nem passa pela minha cabeça, porque eles tiveram uma boa educação. Eu sou um pai presente, então não corro esse risco” (Agência Lupa, 2026, p. 24). Em novembro de 2017, o Tribunal de Justiça do Rio de Janeiro confirmou a condenação do deputado por dano moral coletivo, por declarações com teor homofóbico dadas ao programa; a pena fixada em 2015 previa indenização de R\$ 150 mil, e o processo, na data da checagem, ainda tramitava. “Teor homofóbico” é a descrição da Lupa, e a condenação é do tribunal, segundo a Lupa. Ambas são registradas aqui com essa atribuição, sem classificação acrescentada.

Este capítulo parte de casos como esse: o de um enunciado que uma instância externa classificou, e pergunta o que a Análise do Comportamento pode dizer sobre ele sem repetir a classificação como se fosse explicação. Chamar um enunciado de discurso de ódio, de homofóbico ou de racista descreve o que ele é para quem o classifica. Não diz o que o produz nem o que o mantém. A pergunta do capítulo é a segunda: que contingências mantêm enunciados desse tipo, do lado do falante, da audiência e do grupo a que se referem. Os instrumentos são três: as contingências entrelaçadas (seção 5.1), o mando disfarçado (5.2) e o controle aversivo (5.3). Depois vêm o rótulo do grupo e a transferência de função (5.4), o que os dados permitem afirmar sobre a passagem do discurso ao ato (5.5) e a ficha do Capítulo 3 aplicada ao caso da abertura (5.6).

Uma regra de todo o livro vale aqui com mais força que nos outros capítulos: “discurso de ódio” é rótulo externo. Só se chama assim o enunciado classificado por instância externa, como decisão judicial, remoção por plataforma ou

codificação de estudo científico, com a instância nomeada. Sem classificação externa, descreve-se a função do enunciado e não se o rotula. As definições e as medidas que cada estudo usa são tratadas no Capítulo 1; aqui elas são tomadas como o estudo as traz.

5.1 O que o comportamento do falante produz: aprovação, retirada de ameaças e funções que variam com o contexto

Para saber o que mantém um enunciado, a análise pergunta pelas consequências que ele produz. O que as fontes verificadas permitem é uma lista curta de candidatas, nenhuma testada com falas de políticos.

A primeira é a aprovação ou a atenção de uma audiência. É o reforço social generalizado do Capítulo 3 (seção 3.3): para Guerin (1994), atitudes são enunciados mantidos por esse tipo de reforçamento numa comunidade verbal. Fazzano e Gallo (2015), que propõem uma hipótese funcional para a agressão contra homossexuais (seção 5.3), hipotetizam, além do reforçamento negativo, a existência de reforço social, fornecido pelos pares, por meio da aprovação do grupo. A classe de respostas que eles analisam tem topografias diversas (bater, insultar, torturar, espancar, excluir, discriminar) e é definida pela função, e não pela forma.

A segunda é a remoção ou a redução de estimulação aversiva. Skinner (1957) inclui o escape e a esquiva de estimulação aversiva entre os reforços generalizados, e entre os modos de estimular alguém aversivamente lista muitos tipos de “dano verbal”. Fazzano e Gallo (2015) hipotetizam que a presença de homossexuais seria aversiva para o agressor e que a agressão seria reforçada negativamente e observam que a redução da intensidade do estímulo já basta para o reforçamento, de modo que a violência não precisa terminar em morte para ser reforçada. Que a aversividade seja aprendida e não do estímulo é o ponto deles: não há, no próprio estímulo, algo que seja aversivo e o caso parece ser de aprendizagem. Os processos de aquisição que apontam são a modelagem por reforço diferencial, o seguimento de instruções e a imitação de um modelo, “muito possivelmente” os três juntos.

A terceira é que as funções variam. Mizael (2018), numa revisão da produção analítico-comportamental sobre homossexualidade (seção 5.3), ao listar o que a Análise do Comportamento poderia fazer pelo tema, descreve (citando uma análise contextual do racismo publicada por Guerin em 2005) funções diferentes de “ser racista” que controlariam o comportamento das pessoas: ser aceito, aumentar o status do próprio grupo, evitação, uso impensado de estereótipos.

O que a lista ensina é que o mesmo enunciado pode ter, em ocasiões diferentes ou em falantes diferentes, consequências diferentes, e que nenhuma delas dispensa a pergunta pela ocasião.

Fazzano e Gallo (2015) criticam as explicações que tratam a violência como defesa contra os próprios impulsos do agressor, ou como defesa do ego, porque atribuem a causa a uma entidade “inacessível e imutável”, desconsideram os contextos por que o indivíduo passou e, nas palavras deles, livram agências sociais de agir. Mizael (2018) retoma o ponto ao propor reduzir os essencialismos que tratam a homofobia como entidade dentro do indivíduo. A análise funcional troca a pergunta: do que há dentro de quem diz para o em que condições o dizer ocorre (Skinner, 1953/2003).

CONCEITO EM USO

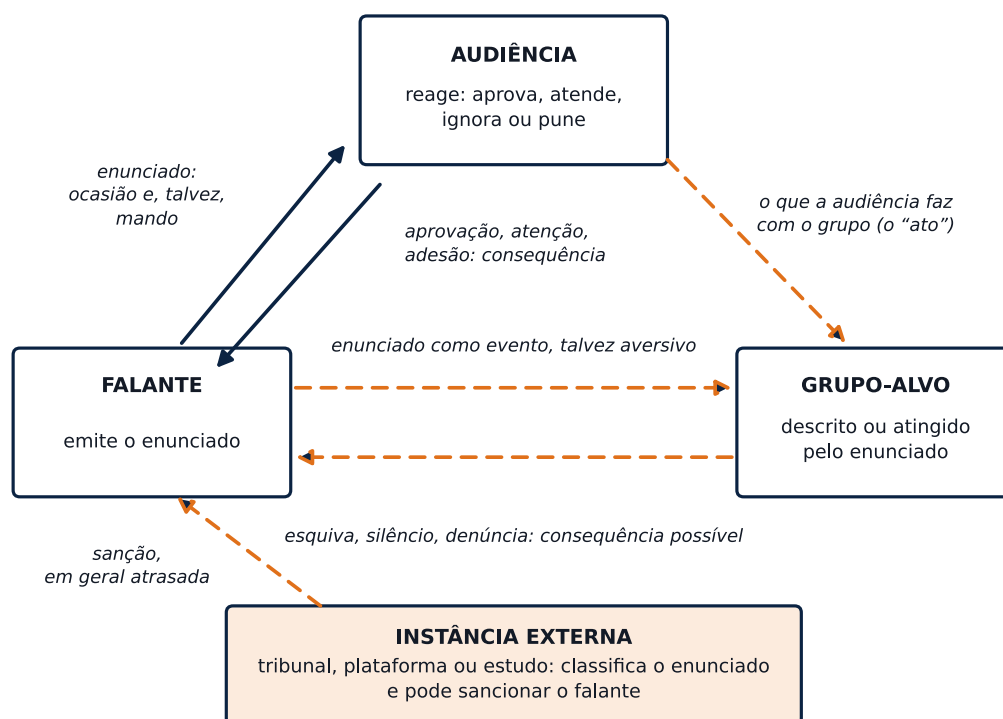
Contingências comportamentais entrelaçadas (Glenn, 2004)

Contingências operantes estão entrelaçadas quando o comportamento de duas ou mais pessoas funciona como evento ambiental para o comportamento das outras (Glenn, 2004). O resultado do comportamento entrelaçado não é o efeito cumulativo de cada pessoa se comportando sozinha, e sim o efeito de seu comportamento inter-relacionado, “mais que” ou “diferente de” o que cada uma produziria isolada. Neste capítulo, o entrelaçamento tem três participantes: o falante, cuja resposta é o enunciado; a audiência, cuja reação é consequência para o comportamento falante; e o grupo-alvo, para quem o enunciado é um evento e cuja reação pode ser também consequência para o comportamento do falante. A aplicação aos três é feita aqui, e não por Glenn.

A Figura 5.1 reúne as relações. As duas linhas contínuas, entre falante e audiência, são as que têm algum dado indireto, tratado na seção 5.5: a fala de líderes que precede mudanças no que seus apoiadores fazem e a expressão pública que acompanha a popularidade percebida do que se diz. As tracejadas são as que o capítulo propõe e nenhuma fonte verificada mede: o que o enunciado faz ao grupo-alvo, o que o grupo-alvo faz de volta, e o que a audiência faz com o grupo. A instância externa não é parte da contingência no momento da emissão: classifica depois e, quando sanciona, o faz com atraso. Só pertencerá a contingência em caso de recorrência e do efeito supressor sobre respostas da mesma classe que a sucederam. No caso da abertura, a consequência documentada para o falante veio de um tribunal, anos depois, e não da audiência do programa.

FIGURA 5.1

Falante, audiência e grupo-alvo: contingências entrelaçadas em torno de um enunciado



Linha contínua: relação com dado indireto em fonte verificada (seção 5.5). Tracejada: relação conceitual ou hipótese.

Nota. Esquema original do autor, a partir de Glenn (2004), Skinner (1957) e Fazzano e Gallo (2015). Não representa dados.

5.2 Mando disfarçado: a opinião que funciona como instrução

O mando, no vocabulário do Capítulo 3, é o operante verbal cuja resposta é reforçada por uma consequência característica e que está, por isso, sob controle de privação ou de estimulação aversiva (Skinner, 1957). Skinner distingue espécies de mando pelas consequências providas ao comportamento do ouvinte ao atender o mando do falante. No pedido, o ouvinte já está disposto a dar o que se pede, e a resposta apenas indica o que o falante aceitará. No comando, a resposta do falante estabelece uma situação aversiva da qual o ouvinte só escapa fornecendo a mediação adequada: “Mãos ao alto!” não só especifica uma ação como constitui uma ameaça da qual a vítima escapa levantando as mãos, e “A bolsa ou a vida!” traz as duas partes, a que especifica o reforço e a que nomeia a consequência aversiva. Skinner escreve que as ordens militares são obedecidas por uma espécie de ameaça permanente. Quando o ouvinte ganha algo próprio, o mando é conselho, se há reforço positivo, ou aviso, se há escape de estimulação aversiva; a permissão cancela uma ameaça que o continha.

CONCEITO EM USO**Mando disfarçado** (Skinner, 1957; Guerin, 1994)

Skinner usa a expressão em duas passagens. Na primeira, a propósito das audiências negativas: o comportamento verbal sobre o ouvinte, ou sobre pessoas importantes para ele, pode ser recebido aversivamente, e com alguns ouvintes passamos a evitar mandos ou a usar mandos disfarçados em seu lugar. Na segunda, o pregão “Call for Philip Morris” funciona como mando embora disfarçado de resposta de um mensageiro que chama alguém. Skinner não desenvolve o conceito além disso. Guerin (1994) o aproxima do que este capítulo precisa: enunciados de atitude podem trazer consequências específicas e, assim, qualificar-se como mandos; a fala da chefe de que não gosta de quem toma café enquanto trabalha não é um tato inocente de atitude privada e funciona como se ela tivesse mandado “Não tome café enquanto trabalha, ou então!”, tenha ou não sido pretendida. Medeiros (2013) define mandos disfarçados como repostas verbais com topografia de tato/intraverbal cuja ocorrência é controlada pelo reforçadores que mantém os mandos, ainda que a sua topografia não explicitamente o reforçador. Neste livro, *mando disfarçado* é o enunciado com forma de tato ou de opinião cuja função, na evidência disponível, é a de mando. A aplicação à fala política é hipótese aqui proposta.

Os dois critérios que o texto de Skinner dá são, portanto, uma condição de aparecimento e um critério de identificação. A condição é uma audiência que pune o mando explícito: com ela, o falante passa a usar a forma que não é punida. O critério é a divergência entre forma e função: a forma é a de outra coisa, a função é a de mando. Este capítulo toma só esses dois e a leitura de Guerin; o resto é aplicação.

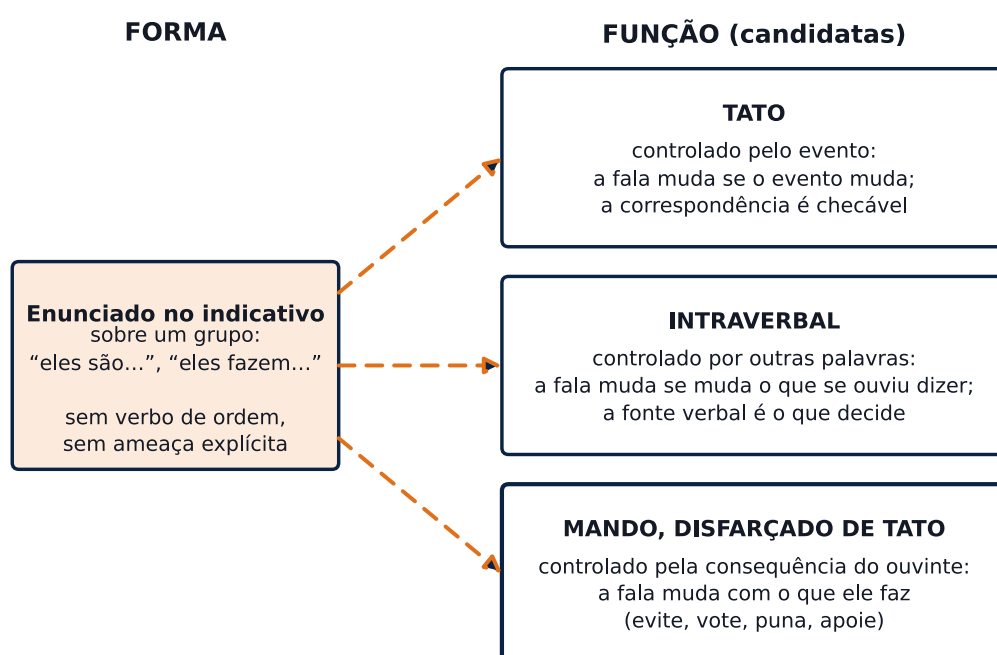
A aplicação é esta. Um enunciado como “eles são perigosos” tem a forma de um tato, no indicativo, sobre um grupo. Se a função for a de mando à audiência, ele será controlado pelo que o ouvinte faz depois, e o que se pede é: evite-os, vote contra, apoie quem os ataca, ao menos não puna. Nada na forma diz isso. A Figura 5.2 mostra as três funções entre as quais a mesma forma pode estar, e que a seção 3.3 já distinguiu: tato, intraverbal e mando, ou seja, os três operantes verbais cuja topografia não é determinada em uma relação ponto-a-ponto com o estímulo antecedente verbal.

O que separa a hipótese de mando de suas rivais são as três perguntas que a ficha do Capítulo 3 já fazia, aplicadas agora ao comportamento do ouvinte. O que o ouvinte faz depois: uma resposta observável e não uma intenção que se lhe atribua. O que o falante faz quando o ouvinte não atende: Skinner nota que a repetição pode ser só evidência de força excessiva da resposta e que, quando o comportamento verbal é ineficaz em alterar as circunstâncias responsáveis

por sua força, uma resposta pode ceder a outra, de modo que, se a repetição pura é punida, a forma varia (seção 3.5). E se a forma muda quando a audiência muda: com uma audiência que pune o mando explícito, o disfarce é esperado; com uma audiência que o reforça, é o mando explícito que se esperaria. Para o autoclítico assertivo, que “ordena ao ouvinte que aceite um dado estado de coisas” e que a seção 3.6 tratou, Skinner já o classifica como tipo especial de mando (p. 326); “nunca”, “jamais” e “sempre” levam, portanto, um mando embutido que não depende do assunto.

FIGURA 5.2

Mando disfarçado: a mesma forma de opinião, três funções candidatas



A forma é a mesma; o que a distingue é o que a controla (Skinner, 1957, p. 36).

Nota. Esquema original do autor, a partir de Skinner (1957) e Guérin (1994). Não representa dados. As três setas são tracejadas porque as três funções são hipóteses para qualquer enunciado particular.

Nenhuma das fontes verificadas registra as três coisas para um enunciado classificado como de ódio. É por isso que a interpretação como mando disfarçado é uma possibilidade a investigar e não uma classificação do enunciado, e é por isso que ela fica, nos quadros do capítulo, na coluna da hipótese.

5.3 Controle aversivo dirigido a um grupo

Skinner (1957) apresenta o controle aversivo do comportamento verbal pela redução da estimulação aversiva. Quando o próprio estímulo aversivo é reduzido, o comportamento é de escape; quando se reduz uma condição que costuma precedê-lo, é de esquiva. Seus exemplos verbais são “Pare!”, reforçado quando faz cessar a injúria física, e “Não me toque!”, reforçado quando faz cessar a ameaça dela. O ponto que importa aqui é o de que a estimulação aversiva não é só física. Skinner escreve que um homem pode estimular outro aversivamente de muitos modos, inclusive por “muitos tipos de dano verbal” (p. 54), e que a ameaça que controla o comportamento verbal não implica necessariamente violência: há formas brandas, e uma pergunta contém uma leve ameaça generalizada, no sentido de que, se não se responde, segue-se censura.

CONCEITO EM USO

Controle aversivo (Skinner, 1957; Fazzano e Gallo, 2015)

Controle aversivo é o controle do comportamento por estimulação aversiva: o comportamento que a remove ou a evita é reforçado (escape e esquiva), e o que a produz tende a ser suprimido (punição). Fazzano e Gallo (2015) definem punidores, com Sidman (1989/2003), como a produção de algo que normalmente seria um reforçador negativo ou a retirada de algo que normalmente seria um reforçador positivo. No caso que analisam, a aversividade não está no estímulo: é aprendida. O recorte que fazem é de uma contingência de três termos, que reconhecem como simplificação das contingências envolvidas nas agressões (nota 3). Estimulação aversiva pode ser verbal (Skinner, 1957) e pode ser branda.

Fazzano e Gallo (2015) são a análise comportamental brasileira da homofobia que as fontes verificadas trazem e ela é, por declaração própria, uma hipótese funcional e não uma análise funcional: faz-se a hipótese, e não a experimentação, por razões éticas. Em seus termos, a homofobia é um conjunto de comportamentos complexos, operantes e respostas emocionais, de agressão física, psicológica ou sexual contra homossexuais ou quem se identifique com a cultura homossexual e se relaciona com o controle aversivo. A parte que interessa a este capítulo é a que trata da manutenção. Para o agressor, as consequências imediatas são a remoção da estimulação aversiva, que as autoras supõem aprendida e, talvez, a aprovação do grupo. As consequências punitivas, como processos penais, multas ou prisões, são distantes no tempo, e o agressor pode nunca entrar em contato com elas, porque, segundo as autoras, em 2015 não havia agravantes para a homofobia no código penal e o número de denúncias registradas era reduzido. O texto é de 2015, e o quadro jurídico pode ter muda-

do; não se afirma aqui o estado atual da lei. Ao mesmo tempo, para que as consideremos como consequências punitivas, é necessário observar a recorrência da relação comportamento consequência. O que não é o caso.

A mesma análise tem um lado da vítima, e é o lado menos trabalhado pelos estudos de discurso. Fazzano e Gallo (2015) tratam a agressão, do ponto de vista de quem a recebe, como punição do comportamento que se emitia: dois homens agredidos por andarem abraçados provavelmente não voltarão a se abraçar no mesmo local ou em contextos parecidos, e a punição generaliza, de modo que a pessoa pode deixar de ir ao lugar da agressão e, em alguns casos, de sair de casa, isolando-se. Permanecer em casa, evitar relacionamentos ou ficar “no armário” seriam respostas incompatíveis, que previnem nova estimulação agressiva. É uma análise dos efeitos de agressões, e não de enunciados e o capítulo a traz porque o enunciado de ódio, se for aversivo para o grupo-alvo, tem esse lado: a consequência que um enunciado desses produz no grupo é uma redução do que o grupo faz em público. Essa análise corresponde ao segundo caso apontado por Skinner, a da adversidade. Se um mando disfarçado resulta em agressões e supressão dos comportamentos dos agredidos, a supressão reforça negativamente o comportamento de emitir o mando disfarçado. Nenhuma fonte verificada mede isso para enunciados.

Há uma ressalva de método das mesmas autoras que vale para todo o livro. O comportamento é definido como violência ou agressão por sua função em determinados contextos: dois amigos que se tratam por um termo geralmente pejorativo podem estar apenas expressando entrosamento e intimidade e o mesmo termo caracterizaria agressão verbal se gerasse na vítima tristeza, constrangimento ou humilhação. Isso concorda com a regra de rótulo deste livro. A palavra não decide: decide o que ela faz, com quem e em que ocasião.

Duas lacunas completam o quadro. A primeira é a que Mizael (2018) documenta numa revisão bibliográfica. A autora buscou os termos “homossexualidade”, “orientação sexual”, “heterossexualidade” e “bissexualidade” (e seus equivalentes em inglês e em espanhol) em 13 periódicos de Análise do Comportamento, três brasileiros e 10 internacionais, sem limite de período e incluiu só os artigos em que a homossexualidade era tema central. Achou três, e uma busca complementar na SciELO, em português, que somou o termo “análise do comportamento”, trouxe o quarto. Um deles é o de Fazzano e Gallo (2015), o único publicado numa revista de psicologia não específica da área. Em geral, os quatro analisavam a homossexualidade ou a homofobia pelas contingências de reforçamento e punição a que as pessoas podem ter sido submetidas ou examinavam possíveis causas biológicas da homossexualidade. Faltavam, segundo a autora, análises funcionais de comportamentos considerados preconceituosos e o estudo de piadas e outros comportamentos potencialmente

discriminatórios como práticas reforçadas pela comunidade verbal. A segunda lacuna é a que Fazzano e Gallo (2015) reconhecem. Os dados brasileiros que citam vêm do Disque 100, canal de denúncias em que a Secretaria de Direitos Humanos se baseou, em 2011, para caracterizar a violência homofóbica: 6.809 relatos de violações contra lésbicas, gays, bissexuais e transexuais, envolvendo 1.713 vítimas e 2.275 suspeitos. Em 61,9% dos casos as vítimas conheciam o agressor, e as formas de violência mais frequentes foram a psicológica (42,5%), a discriminação (22,3%) e a violência física (15,9%). Como são só denúncias, as próprias autoras escrevem que os números podem não expressar significativamente a realidade brasileira.

5.4 O rótulo do grupo e a transferência de função aversiva

O Capítulo 4 mostrou como uma palavra adquire função avaliativa sem contato direto com o referente: por pertencer a uma classe de equivalência ou por entrar em relações de oposição e de comparação com outras (seções 4.2, 4.3 e 4.6). Para o discurso sobre grupos, isso quer dizer que o enunciado não precisa de contato prévio direto do ouvinte com o grupo para que uma função aversiva se transfira ao grupo; basta que o rótulo do grupo esteja numa classe ou numa relação com estímulos de função aversiva. Fazzano e Gallo (2015) têm o ponto em outros termos: se a presença de homossexuais não é aversiva por si, a aversividade foi aprendida, e a aprendizagem pode ocorrer por modelagem, por seguimento de instruções e por imitação, e o seguimento de instruções é um processo verbal. Caso uma frase com autoclítico assertivo adjective de forma pejorativa um grupo (“Mulheres votam mal!”), mesmo que o ouvinte não tenha entrado em contato com estímulos aversivos decorrentes do voto de mulheres de forma direta, a possibilidade de mulheres votarem adquire função aversiva por transferência de função ou transitividade.

Mizael (2018) dá um exemplo do que a transferência quer dizer para o preconceito sexual, o dá como ilustração teórica e proposta de pesquisa, e não como dado: um menino pode aprender que “viado” é algo ruim sem saber o significado da palavra; quando cresce, aprende que a palavra é equivalente a “gay” ou “homossexual”, e os atributos negativos passam a quem se define ou é percebido como gay. O mecanismo que o exemplo descreve é o que o Capítulo 4 mostrou em laboratório com outros estímulos. Dixon et al. (2003), no mesmo sentido, analisam preconceito e atos de terror como produto da transformação de funções de indivíduos ou grupos com base em contato frequentemente indireto com poucos de seus membros, como argumento e sem dados (Capítulo 4, seção 4.6).

O que isso acrescenta ao capítulo é uma terceira posição para o falante, além de objeto e agente de aversão: a de quem estabelece as relações. Um enunciado que associa o rótulo de um grupo a termos de função aversiva, ou que o opõe a um “nós” de função positiva, estaria fazendo, para o ouvinte, o trabalho que o procedimento de laboratório faz para o participante. Dado, aqui, só há o procedimento; a hipótese de que a fala política o faça, e com que efeito sobre o que o ouvinte diz e faz depois, é aqui proposta, e para testá-la seriam necessários falas datadas e medidas do ouvinte antes e depois, e não só a análise do texto.

5.5 Do discurso ao ato: o que os dados permitem afirmar

A passagem do discurso ao ato junta duas perguntas que convém separar. A primeira é se o enunciado muda o que a audiência faz, por exemplo agir contra o grupo ou deixar de punir quem age. A segunda é se o enunciado muda o que o próprio falante ou outros falantes dizem em seguida. Os estudos verificados são, em sua maior parte, da segunda; os da primeira não tratam de discurso de ódio. A Tabela 5.1 os reúne.

TABELA 5.1

O que seis estudos verificados medem como “discurso”, como “ato” e o que permitem afirmar

Estudo	Medida de “discurso”	Medida de “ato” ou de efeito	O que permite	O que não permite
Ajzenman et al. (2023, versão de trabalho)	Participação do presidente em manifestação (15/3/2020) e pronunciamento (24/3/2020) contra o distanciamento social; não é discurso sobre grupo	Índice de distanciamento por celulares, por município	Queda do distanciamento maior em municípios onde Bolsonaro teve mais de 50% dos votos no primeiro turno de 2018, cerca de 1 ponto percentual em média, sem efeitos antes dos eventos	Separar conduta de fala; falar de discurso de ódio; mecanismo (mídia, evangélicos) é só sugestivo
Bursztyn et al. (2020)	Informação aleatória sobre quem ganhou na área (popularidade percebida de Trump)	Doação a organização anti-imigração, em público ou em privado	A doação pública, e não a privada, variou com a popularidade percebida	Efeito de falas; só manipula a percepção de popularidade; EUA

Estudo	Medida de “discurso”	Medida de “ato” ou de efeito	O que permite	O que não permite
Álvarez-Benjumea e Winter (2018)	Comentários em fórum experimental, com nota de hostilidade	O próprio comentário	Remover o conteúdo de ódio anterior reduziu comentários hostis (5,88% para 0,99%); a contra-fala não teve efeito significativo	Generalizar; participantes pagos e anônimos
Siegel et al. (2021)	Tweets classificados por dicionário e por comunidades do Reddit	Nível agregado do discurso, sem medida de ato	Sem aumento persistente de discurso de ódio na campanha nem depois da eleição de 2016	Excluir outras plataformas; dicionário conservador
Marino-Fages e Martínez-Pozo (2026, documento de trabalho)	Tweets geolocalizados classificados por modelo e por dicionário	Nível do discurso, por município	Aumento depois da eleição de 2018, maior onde Bolsonaro teve menos votos, em homofobia e sexismo	Medir custo social percebido; o mecanismo é leitura dos autores
Hangartner et al. (2021)	Tuítes xenófobos de 1.350 usuários	Exclusão retrospectiva e criação prospectiva de discurso de ódio	Contra-fala baseada em empatia aumentou a exclusão e reduziu a criação; humor e aviso de consequências, sem efeito consistente	Efeito sobre quem apenas lê

Nota. Números dos artigos. Ajzenman et al. foram lidos na versão de trabalho de 2020, da Universidad Torcuato Di Tella, sem paginação do periódico. Marino-Fages e Martínez-Pozo é documento de trabalho que, segundo a capa, não passou por revisão de pares. Os estudos de Bursztyn et al., de Álvarez-Benjumea e Winter, de Siegel et al. e de Marino-Fages e Martínez-Pozo são descritos no Capítulo 10 (seções 10.3, 10.4 e 10.6); o de Hangartner et al. é tratado no Capítulo 12.

O estudo de Ajzenman et al. (2023) é o único, entre os verificados, que liga a fala de um chefe de Estado brasileiro ao que sua audiência faz e o que mede é distanciamento durante a pandemia e não ato contra um grupo. A hipótese dos autores era que os apoiadores do presidente seriam bem mais propensos a se deixar persuadir por seus discursos do que seus críticos, mesmo em decisões de alto custo, como seguir ou não as medidas preventivas recomendadas pela Organização Mundial da Saúde e pelos governos subnacionais. O desfecho foi um índice de distanciamento calculado a cada dia, para cada município, a partir da localização de cerca de 60 milhões de celulares anônimos: a proporção dos aparelhos que ficou, no dia, a menos de 450 metros do local identificado como a casa. O índice vai de 0 a 1, teve média de 0,37 no período (de fevereiro às duas primeiras semanas de abril de 2020) e existe para 3.975 dos 5.571 municípios, porque os pequenos não têm aparelhos suficientes (seção II). Os dois eventos estudados foram a participação do presidente nas ma-

nifestações de 15 de março de 2020 e o pronunciamento oficial de 24 de março, em que pressionou pelo fim das medidas de distanciamento dos governos sub-nacionais.

A comparação é entre os municípios pró-governo, em que Bolsonaro teve mais de 50% dos votos no primeiro turno de 2018 e os demais, dia a dia, antes e depois dos dois eventos. Os autores chamam o desenho de estudo de eventos, com efeitos fixos de dia e de município (que descontam o que é próprio de cada município e o que muda em todos no mesmo dia) e estimativas separadas para cada dia anterior (adiantamentos) e posterior (defasagens) ao evento (seções I e V.1). Nos dias anteriores, com uma exceção, a diferença entre os dois grupos não difere de zero, isto é, não havia diferença prévia entre eles. A partir do segundo dia depois dos eventos, o distanciamento foi significativamente menor nos municípios pró-governo, cerca de 1 ponto percentual em média. Numa especificação que se restringe aos estados em que ele teve mais de 50% dos votos em ao menos um terço dos municípios e compara, em cada estado, os municípios acima e abaixo da mediana de votos, o efeito fica perto de 2 pontos percentuais, frente a uma média de 0,37 no período (V.2). Com os coeficientes e a população dos municípios pró-governo (cerca de 100 milhões), os autores calculam, em conta aproximada, cerca de 1 milhão de pessoas a mais afastadas mais de 450 metros de casa em cada um dos dez dias seguintes. Chamam o resultado de “forte efeito de persuasão”, e dizem não poder apresentar resultados conclusivos sobre os mecanismos (V.3 e VI).

Em termos operantes, o estudo é compatível com uma leitura do pronunciamento como regra, ou como mando, para quem já tinha história de reforço com o presidente e é compatível também com a imitação da conduta dele na manifestação. Os autores estimam o efeito conjunto da participação nas manifestações e das mensagens contra o distanciamento (seção V.1), e, portanto, não distinguem entre as duas. O ganho para este capítulo é de método: medir o que a audiência faz depois de uma fala é viável em larga escala para condutas que os celulares registram. Já a pergunta pelo ato contra um grupo exigiria outra medida. A condição de leitura é que o efeito aparece onde a audiência já apoiava o líder. Baldwin e Baldwin (1986) sustentam aspectos da relação do observador com o modelo (admiração, respeito e afeto) como elementos fundamentais na probabilidade do observador se comportar como o modelo. Não é dado sobre discurso de ódio, e não deve ser lido assim.

Os quatro estudos seguintes, descritos com mais vagar no Capítulo 10, tratam do que as pessoas dizem ou aceitam fazer em público depois de uma informação ou de um evento e dois deles são experimentos. Bursztyn et al. (2020) recrutaram 1.600 moradores da região de Pittsburgh num painel online e informaram, ao acaso, metade deles de que Trump ganhara na região metropolita-

na e a outra metade de que Clinton ganhara no condado onde fica a cidade. Ofereceram então um bônus a quem autorizasse uma doação a uma organização fortemente anti-imigração e a decisão era anônima (condição privada) ou visível a pares da região (condição pública). Em público, aceitaram doar 33% de quem soube da vitória de Trump e 20% de quem soube da de Clinton; em privado, 31% e 30%. A diferença entre essas diferenças foi significativa: o efeito da informação apareceu em público e não em privado.

Álvarez-Benjumea e Winter (2018) recrutaram 180 participantes numa plataforma de crowdsourcing e pediram que comentassem fatos de um fórum experimental, em quatro condições: a linha base, com comentários amistosos, neutros e hostis; a censurada, com o conteúdo de ódio anterior apagado; a extremamente censurada, só com comentários amistosos; e a de contra-fala, com os comentários hostis acompanhados de réplicas que destacavam a inaceitabilidade. Os comentários hostis foram 5,88% na linha de base e 0,99% na condição censurada e só essa redução foi significativa; a contra-fala (4,44%) não teve efeito significativo.

Os outros dois são estudos observacionais de textos em redes sociais. Siegel et al. (2021) analisaram mais de 750 milhões de tweets sobre a eleição americana de 2016 e quase 400 milhões de uma amostra aleatória de usuários americanos e com as duas medidas que adotaram (um dicionário aumentado por aprendizado de máquina e uma classificação baseada em comunidades do Reddit ligadas à direita alternativa), não acharam aumento persistente de discurso de ódio na campanha nem depois da eleição; houve só picos breves.

Marino-Fages e Martínez-Pozo (2026) reuniram mais de 37,6 milhões de tweets em português, geolocalizados no Brasil, de julho de 2017 a dezembro de 2019, e compararam o antes e o depois da eleição de 2018 segundo a parcela de votos de Bolsonaro no primeiro turno em cada município: o discurso de ódio aumentou depois da eleição, mais onde ele teve menos votos, e o aumento se concentrou em homofobia e sexismo. As quatro medidas e os quatro desenhos diferem e os estudos não se confirmam nem se refutam. Hangartner et al. (2021), no Capítulo 12, completam a lista do lado da intervenção: sortearam usuários do Twitter autores de tuítes xenófobos para receber ou não uma contra-fala e a de empatia aumentou a exclusão do tuíte original e reduziu a criação de novos tuítes xenófobos com ressalvas que o Capítulo 12 detalha.

Álvarez-Benjumea e Winter (2018) tiram do experimento uma conclusão que interessa à leitura operante: o comportamento normativo online seria motivado mais por normas descritivas - o que os outros fazem - do que por normas injuntivas - o que se deve fazer. No vocabulário usado aqui, a norma descritiva é o que a audiência faz diante do enunciado, estímulo discriminativo para

quem vai falar em seguida; a injuntiva é o que a comunidade diz que se deve fazer, uma regra (seção 3.7). A condição de contra-fala traz a injuntiva, com réplicas que dizem ser inaceitável o ódio e não teve efeito significativo; a condição censurada muda o que se vê os outros dizerem e teve efeito. Os autores sugerem, para o fracasso da contra-fala, que réplicas, elas mesmas hostis, acrescentam ambiguidade sobre a norma. O experimento não separa as duas leituras e é um experimento só com participantes pagos e anônimos; a leitura é compatível com a de que, para quem fala, o que a audiência faz pesa mais do que o que ela diz que se deve fazer.

Falta, na Tabela 5.1, a coluna que o título da seção pedia: a do ato contra o grupo. Entre as fontes verificadas, nenhuma liga enunciados de ódio a atos de violência contra o grupo-alvo no Brasil ou fora dele. Estudos que o fazem existem na literatura e não são citados aqui porque o texto integral não pôde ser lido por via legal. A lacuna é, portanto, do acervo de fontes lidas e deve ser lida como tal: não é ausência de relação, nem sua demonstração.

<p>■ DADO</p>	<p>Em municípios onde o presidente teve mais votos, o distanciamento social caiu depois de dois eventos em que ele se opôs publicamente ao distanciamento, cerca de 1 ponto percentual em média, sem efeito antes dos eventos (Ajzenman et al., 2023). A expressão pública de posição anti-imigração e não a privada acompanhou a popularidade percebida do candidato (Bursztyn et al., 2020). No Brasil, o discurso de ódio em tweets geolocalizados aumentou depois da eleição de 2018, mais onde Bolsonaro teve menos votos (Marino-Fages e Martínez-Pozo, 2026).</p>
<p>■ INTERPRETAÇÃO</p>	<p>O que a audiência faz e o que os outros dizem parecem depender do que a audiência acredita que os outros aceitam e do que o líder faz e diz. As evidências são indiretas, de desenhos e medidas diferentes, e nenhuma mede a fala de um político de modo a poder atribuir o efeito ao conteúdo e não à conduta, ao cargo ou ao momento.</p>
<p>■ HIPÓTESE</p>	<p>Que o enunciado de ódio de um político funcione como mando à audiência, com efeito mensurável sobre o que ela faz com o grupo-alvo é hipótese, e nenhuma fonte a testa. Testá-la exigiria falas datadas, medida do que a audiência faz depois e comparação com quem não foi exposto e, para o grupo-alvo, medidas do que ele deixa de fazer em público.</p>

5.6 Um caso com classificação externa: a fala de 2011 e o que veio depois

O caso da abertura é o único entre as fontes verificadas com os três elementos que a regra do capítulo pede para falar em discurso de ódio sem rotular por conta própria: o enunciado transcrito, a data e o veículo, e a classificação por instância externa, nomeada e datada. O que não tem é o que a ficha do Capítulo 3 pede em audiência presente e em variáveis de controle candidatas, a composição da audiência e a reação dela, que ficam em branco; a seção diz quais.

Os campos documentados são estes. A fala: “Isso nem passa pela minha cabeça, porque eles tiveram uma boa educação. Eu sou um pai presente, então não corro esse risco”, em resposta a uma pergunta sobre o que faria se tivesse um filho gay (Agência Lupa, 2026). A data e o veículo: 2011, programa CQC, da TV Bandeirantes; o mês não consta da fonte. A classificação externa: condenação por dano moral coletivo, confirmada pelo TJRJ em novembro de 2017, depois de declarações com teor homofóbico ao programa, com pena fixada em 2015 de R\$ 150 mil, tudo segundo a Lupa; o texto do tribunal não foi lido. Ainda segundo a Lupa, o processo seguia em tramitação em 2018 e há outras declarações do deputado com teor que ela chama de homofóbico, em debate na TV Câmara e em entrevista à *Playboy*, em 2011. Em 3 de agosto de 2018, na mesma checagem, a frase “Eu nunca fui homofóbico”, dita ao Central das Eleições, da GloboNews, foi classificada pela Lupa como “Falso” (Capítulo 3, seção 3.8, caso 4).

Os campos da ficha que ficam em branco são os mais importantes para a análise operante: a reação da audiência do programa, a composição dela, o que o falante fez depois com audiências diferentes, e qualquer medida do que o grupo-alvo fez ou deixou de fazer. Nada disso está nas fontes e o capítulo não os preenche por inferência. A negação de 2018 foi classificada pela Lupa e figura entre os casos do Capítulo 3, onde recebeu o mesmo tratamento que as falas de Lula; a análise lá é da forma da frase e da correspondência com o dado, e a daqui, da relação entre a fala de 2011, a sanção e a fala de 2018.

Uma ressalva de simetria, no mesmo espírito da que o Capítulo 3 fez. As fontes verificadas não trazem, para outro campo político, caso equivalente, isto é, com enunciado transcrito, datado e classificado por instância externa nomeada. Isso é limite do acervo de fontes lidas e não propriedade dos campos políticos. A ficha e o critério seriam os mesmos, e a regra de amostra da seção 2.9 é o que permitiria encontrar os casos.

■ DADO	Uma resposta, dada em 2011 num programa de TV a uma pergunta sobre o que o deputado faria se tivesse um filho gay, levou a uma condenação por dano moral coletivo confirmada em 2017, segundo a Lupa. Em 2018, o mesmo político, em entrevista, disse nunca ter sido homofóbico, e a Lupa classificou a frase como “Falso”.
■ INTERPRETAÇÃO	A resposta foi emitida diante de uma pergunta, e o enunciado se apoia, na forma, em duas descrições do próprio falante (“eu sou um pai presente”, “eles tiveram uma boa educação”) e numa palavra de função avaliativa (“risco”), que faz de ter um filho gay uma ocasião a evitar. Só pela forma, não se decide se o enunciado é um tato sobre o que o falante faria, um intraverbal aprendido ou um mando à audiência (seções 3.3 e 5.2). A condenação é uma consequência aversiva, atrasada em anos, e vem de uma instância jurídica e não da audiência do programa. A condenação por dano moral coletivo é uma classificação jurídica do efeito sobre o grupo; não é uma medida do comportamento do grupo-alvo.
■ HIPÓTESE	Que a fala de 2011 tenha sido reforçada pela audiência do programa (atenção, aprovação, adesão) é hipótese, e a fonte nada registra sobre a reação dela. Que a condenação tenha alterado a emissão de enunciados desse tipo e que a negação de 2018 (“Eu nunca fui homofóbico”) seja função dessa história, seja como esquiva de punição, seja como tato distorcido pela consequência do ouvinte (seção 3.4) é hipótese e exigiria uma série de falas datadas, antes e depois de 2015 e de 2017, com as audiências.

5.7 Interpretações rivais e limites

A leitura do capítulo, de que o enunciado de ódio é mantido por consequências sociais e aversivas e pode funcionar como mando disfarçado, tem rivais, e a Tabela 5.2 as reúne com o que as separaria dela.

Os limites do capítulo são os das fontes. Sobre agressão e discurso de ódio, o acervo de fontes lidas traz um artigo de hipótese funcional sobre homofobia (Fazzano & Gallo, 2015) e uma revisão que mostra a escassez da área (Mizael, 2018); os experimentos do Capítulo 4 tratam de preconceito, não de enunciados de ódio, não há estudo que teste a função de mando ou o controle aversivo para enunciados. As evidências sobre o discurso e o ato são de desenhos e medidas diversos de um país, de uma plataforma ou de um fórum experimental e nenhuma é de fala de político com audiência identificada. O único caso com classificação externa é de um campo político só e a classificação é de uma ins-

tância e dos relatos da Lupa, registrados com a atribuição. As relações tracejadas da Figura 5.1 não têm medida entre as fontes. A seção 5.5 diz quais lacunas são do acervo de fontes lidas, e não da literatura.

TABELA 5.2

Interpretações rivais para um enunciado classificado como de ódio e o que as separaria da leitura operante principal

Rival	O que diz	O que a separaria
Estado emocional	O enunciado é efeito de um estado de raiva ou de aversão, mais do que de consequência social; Skinner nota que provocar raiva aumenta a probabilidade de comportamento abusivo ou agressivo (1957)	Se a emissão varia com a presença, a ausência e a identidade da audiência; se cessa quando não há ninguém
Identidade e ritual de grupo	A troca é o reforço, e o conteúdo varia sem alterar o efeito (Guerin, 1994)	Se o conteúdo muda sem mudar a reação; se a reação é à adesão e não ao dito
Intraverbal	Repete o que se ouviu na família, na religião ou no grupo, por modelação e regra (Fazzano e Gallo, 2015)	História de aquisição; se varia com a fonte verbal; se é sensível ou insensível a consequências
Brincadeira	A função é entrosamento ou atenção e não agressão (Fazzano & Gallo, 2015; Mizael, 2018)	O efeito sobre o grupo-alvo e sobre a audiência; se o termo gera constrangimento ou humilhação em quem é descrito
Tato sob controle de evento	A frase é relato de evento, com correspondência checável	Checagem da correspondência por agência; se muda com o evento
Artefato da classificação	Outra instância, com outro critério, não a classificaria assim; reclassificar uma amostra por (os classificadores e dicionários usados nos estudos erram: acurácia de 77 a 79% em Marino-Fages e Martínez-Pozo, 2026)	Comparar critérios e instâncias;

Síntese

Enunciado classificado como de ódio é um rótulo de instância externa e não uma explicação. A análise operante pergunta que consequências o mantêm, que função tem para a audiência e que efeito tem para o grupo a que se refere. Para o falante, as candidatas são a aprovação da audiência, a remoção de estimulação aversiva aprendida e funções que variam com a ocasião. Para a audiência, a hipótese é a de que um enunciado com forma de tato opere como mando disfarçado: condição de aparecimento, audiência que pune o mando

explícito; critério, forma de uma coisa, função de outra. Para o grupo-alvo, a estimulação é aversiva e, na análise de Fazzano e Gallo (2015) para a agressão, pune o que o grupo fazia em público e reforça o que o esconde.

Os dados verificados não ligam enunciado de ódio a ato contra o grupo. Mostram, com medidas e desenhos que não se somam, que a fala de líderes precede mudanças no que seus apoiadores fazem (no distanciamento social, não em ódio), que a expressão pública acompanha a popularidade percebida, que a remoção moderada de conteúdo hostil reduz a hostilidade num fórum experimental e que, no Brasil, o discurso de ódio em tweets aumentou depois da eleição de 2018, mais onde o vencedor teve menos votos. O único caso com classificação externa nas fontes é a fala de 2011 e a condenação confirmada em 2017, segundo a Lupa; os campos da ficha que importam à análise operante estão em branco, e ficam assim.

O Capítulo 6 retoma o controle aversivo pelo lado de quem o recebe, como contracontrole (seção 6.4); o 10 define a janela de uma audiência pelo que ela pune e pelo que não pune e descreve com mais vagar os estudos da Tabela 5.1; o 12 trata da contra-fala e das demais intervenções sobre o falante e sobre a audiência.

Referências

- Agência Lupa. (2026, 27 de julho). *Anatomia da desinformação eleitoral* [Relatório]. https://uploads.agencialupa.org/2026/07/RELATORIO-ANATOMIA-DA-DESINFORMACAO-ELEITORAL_LUPA-FINAL.pdf
- Ajzenman, N., Cavalcanti, T., & Da Mata, D. (2023). More than words: Leaders' speech and risky behavior during a pandemic. *American Economic Journal: Economic Policy*, 15(3), 351–371. <https://doi.org/10.1257/pol.20210284>
- Álvarez-Benjumea, A., & Winter, F. (2018). Normative change and culture of hate: An experiment in online environments. *European Sociological Review*, 34(3), 223–237. <https://doi.org/10.1093/esr/jcy005>
- Baldwin, J. D., & Baldwin, J. I. (1986). *Behavior Principles in Everyday Life*. New Jersey: Prentice-Hall.
- Bursztyn, L., Egorov, G., & Fiorin, S. (2020). From extreme to mainstream: The erosion of social norms. *American Economic Review*, 110(11), 3522–3548. <https://doi.org/10.1257/aer.20171175>
- Dixon, M. R., Dymond, S., Rehfeldt, R. A., Roche, B., & Zlomke, K. R. (2003). Terrorism and relational frame theory. *Behavior and Social Issues*, 12(2), 129–147. <https://doi.org/10.5210/bsi.v12i2.40>
- Fazzano, L. H., & Gallo, A. E. (2015). Uma análise da homofobia sob a perspectiva da Análise do Comportamento. *Temas em Psicologia*, 23(3), 535–545. <https://doi.org/10.9788/tp2015.3-02>

- Glenn, S. S. (2004). Individual behavior, culture, and social change. *The Behavior Analyst*, 27(2), 133–151. <https://doi.org/10.1007/bf03393175>
- Guerin, B. (1994). Attitudes and beliefs as verbal behavior. *The Behavior Analyst*, 17(1), 155–163. <https://doi.org/10.1007/bf03392661>
- Hangartner, D., Gennaro, G., Alasiri, S., Bahrich, N., Bornhott, A., Boucher, J., Demirci, B. B., Derksen, L., Hall, A., Jochum, M., Munoz, M. M., Richter, M., Vogel, F., Wittwer, S., Wüthrich, F., Gilardi, F., & Donnay, K. (2021). Empathy-based counterspeech can reduce racist hate speech in a social media field experiment. *Proceedings of the National Academy of Sciences*, 118(50), Article e2116310118. <https://doi.org/10.1073/pnas.2116310118>
- Marino-Fages, D., & Martínez-Pozo, A. (2026). *Hate in the Tropics: Political leaders and the social acceptability of online hate speech* (Documento de trabajo N.º 391). RedNIE. <https://rednie.eco.unc.edu.ar/files/DT/391.pdf>
- Medeiros, C. A. (2013). Mentiras, Indiretas, Desculpas e Racionalizações: Manipulações e Imprecisões do Comportamento Verbal. Associação Brasileira de Psicologia e Medicina Comportamental: *Comportamento em Foco* v. 2 (pp. 157- 170). São Paulo: ABPMC.
- Mizael, T. M. (2018). Perspectivas analítico-comportamentais sobre a homossexualidade: análise da produção científica. *Revista Perspectivas*, 9(1), 15–28. <https://doi.org/10.18761/PAC.2017.011>
- Siegel, A. A., Nikitin, E., Barberá, P., Sterling, J., Pullen, B., Bonneau, R., Nagler, J., & Tucker, J. A. (2021). Trumping hate on Twitter? Online hate speech in the 2016 U.S. election campaign and its aftermath. *Quarterly Journal of Political Science*, 16(1), 71–104. <https://doi.org/10.1561/100.00019045>
- Skinner, B. F. (2003). *Ciência e comportamento humano*. J. C. Todorov & R. Azzi (Trad.). V. 10. Martins Fontes.
- Skinner, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts.



PRÁTICA BASEADA EM EVIDÊNCIAS

Como uma prática vira evidência, e o que a evidência não diz

O módulo completo sobre prática baseada em evidências: delineamentos, critérios, níveis de evidência e os limites de cada um. A mesma pergunta que este livro faz aos dados, feita à clínica.

Acesse, de graça e sem cadastro

walden4.com.br/pratica-baseada-em-evidencias



6

Bolhas digitais como comunidades verbais e a esquiva de contracontrole

Nyhan et al. (2023) testaram, num experimento de campo no Facebook durante a eleição presidencial americana de 2020, o que acontece com as atitudes quando se reduz a exposição a fontes de mesma inclinação política. Participaram 23.377 usuários nos Estados Unidos. Para os designados ao tratamento, os autores rebaixaram no feed, de 24 de setembro a 23 de dezembro de 2020, o conteúdo de amigos, grupos e páginas que se previa compartilharem a inclinação política do participante; o feed do grupo de controle não foi alterado de modo sistemático. A parcela do que os usuários viam vinda dessas fontes caiu de 53,7% (controle) para 36,2% (tratamento), cerca de um terço a menos. Os autores mediram, então, oito atitudes que haviam registrado de antemão (pré-registro), como polarização afetiva, extremismo ideológico, avaliação de candidatos e crença em alegações falsas, e não encontraram efeito mensurável em nenhuma: descartam efeitos de 0,12 desvio-padrão ou mais, isto é, de pouco mais de um décimo da variação típica de cada medida. Os dados de exposição do mesmo artigo, de 2020, que cobrem a população adulta ativa do Facebook nos Estados Unidos, dizem que o conteúdo de fontes de mesma inclinação é a maior parte do que os usuários veem. Se a exposição homogênea fosse o que sustenta as atitudes, cortar um terço dela deveria mover alguma coisa, e não moveu.

O resultado tem duas leituras. Na primeira, a bolha não tem o efeito que se lhe atribui. Na segunda, o que se reduziu não era o que importa: a literatura mede a bolha pela exposição, isto é, pelo que chega ao feed, e a análise do comportamento verbal pergunta por outra coisa, por quem reforça o que o falante diz e por quanto do que vem de fora da comunidade que o reforça chega até ele.

Este capítulo toma a segunda leitura como hipótese de trabalho e diz, a cada passo, o que a sustenta. A primeira será apresentada, com seus argumentos, na seção 6.8.

O percurso tem três partes. As seções 6.1 a 6.3 descrevem a bolha como comunidade verbal e a confrontam com os dados sobre câmaras de eco. A seção 6.4 trata do que protege o falante da correção, que é a proposta central do capítulo e a parte de menor lastro empírico. As seções 6.5 e 6.7 a levam ao material brasileiro e a um método do autor. Em todas, dado, interpretação e hipótese ficam separados, e chamar um repertório de esquiva de contracontrole é sempre interpretação.

6.1 A bolha como comunidade verbal: audiência e fonte de reforço

Guerin (1992) lembra que a comunidade verbal de Skinner, em escala menor, equivale aos coletivos de Durkheim e aos outros generalizados de Mead, e que mudanças no uso da linguagem podem estar sob controle de subcomunidades verbais menores. É nesse sentido que este capítulo fala de bolha: não um lugar nem uma distribuição de conteúdos, mas um conjunto de ouvintes com história de reforçamento comum, que reforça certas coisas ditas e outras não.

CONCEITO EM USO

Subcomunidade verbal (Guerin, 1992), retomada do Capítulo 3

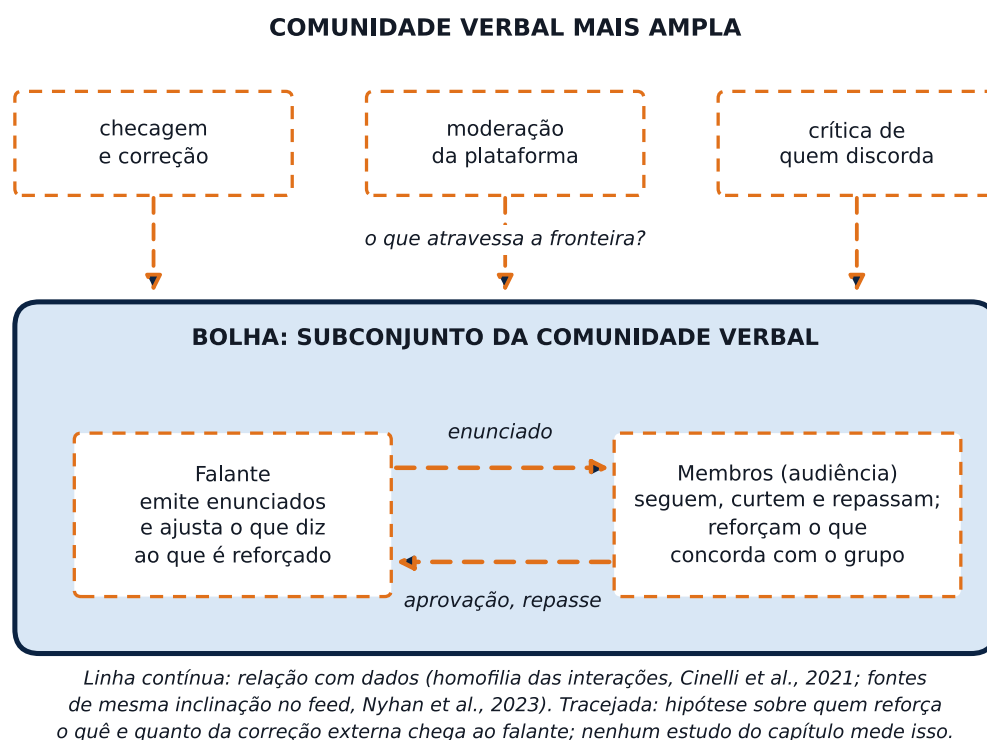
O Capítulo 3 (seção 3.4) apresentou o tato distorcido: o tato cujo controle pelo evento foi deslocado por consequências especiais do ouvinte (Skinner, 1957). O que este capítulo acrescenta é a escala. Para Guerin (1992), o controle do que se diz pode estar nas mãos de subcomunidades verbais menores que a comunidade verbal ampla: se um pequeno grupo reforça certos relatos e pune outros, as características do grupo passam a controlar o que se relata, ainda que as consequências sejam generalizadas, e Skinner (1957, citado por Guerin, 1992) chamou esses tatos de distorcidos. A seção 6.3 traz as três vias pelas quais, segundo Guerin, o que se diz se desliga do ambiente físico e passa a ser controlado pelo ambiente social.

O Capítulo 3 apresentou os dois termos de que a descrição precisa: a audiência é ocasião para o que se diz (seção 3.1), e a reação dela é consequência (seção 3.3; Guerin, 1994). Falta dizer o que muda quando a audiência é uma subcomunidade. Muda a composição das reações. Se os ouvintes que reagem compartilham uma história, as reações que o falante recebe tendem a vir da mesma direção, e o que a comunidade mais ampla diria do enunciado, corrigir, criticar,

não aderir, tem pouco efeito ou chega por outras vias. A hipótese do capítulo é que isso, e não a mera exposição, é o que faz da bolha um ambiente de controle verbal distinto.

FIGURA 6.1

A bolha como subcomunidade verbal: o que os estudos medem e o que o modelo supõe



Nota. Esquema original do autor, sem dados. Linha contínua: relação com dados (homofilia das interações, Cinelli et al., 2021; predominância de fontes de mesma inclinação no feed, Nyhan et al., 2023). Tracejada: hipótese. As relações internas à bolha (quem reforça o quê) e o que da correção externa chega ao falante não foram medidos em nenhum estudo do capítulo.

A literatura mostra homofilia nas interações e predominância de fontes de mesma inclinação no feed (seção 6.2). As setas internas, em que o falante ajusta o que diz ao que é reforçado e a audiência reforça o que concorda com o grupo, têm apoio na análise de Guerin e, para opiniões, nos experimentos de 1956 a 1969 do Capítulo 9, mas não foram medidas em bolhas digitais. E a pergunta que a figura faz na fronteira, quanto da correção dirigida ao que o falante disse chega a ele, tampouco foi medida: Nyhan et al. (2023) medem a parcela de conteúdo de fontes cruzadas no feed, que não é a mesma coisa.

6.2 O que os dados dizem sobre câmaras de eco (e onde divergem)

A literatura usa “câmara de eco” de modos diferentes, e a escolha da medida impacta na resposta observada. Cinelli et al. (2021) definem câmara de eco como ambiente em que a opinião, a inclinação política ou a crença do usuário sobre um tema é reforçada por interações repetidas com pares ou fontes de tendências e atitudes semelhantes, e a operacionalizam por dois dados observáveis, a homofilia nas interações sobre um tema (usuários que interagem com quem pensa como eles) e o viés da difusão da informação a partir de pares de mesma inclinação, que devem ser quantificados independentemente. Analisaram mais de 100 milhões de conteúdos sobre temas controversos (como controle de armas, vacinação e aborto) em quatro plataformas, Gab, Facebook, Reddit e Twitter, com um conjunto de dados por plataforma: aborto (Twitter), vacinas (Facebook), política (Reddit) e o Gab inteiro. A inclinação dos usuários, um valor de -1 a +1 sobre o tema, foi inferida de curtidas em páginas rotuladas (no Facebook, por exemplo, antivacina vale +1 e provacina, -1), de links para veículos classificados por Media Bias/Fact Check (2.190 veículos) e de comentários. A homofilia foi medida pela correlação entre a inclinação de cada usuário e a média da inclinação de seus vizinhos na rede de interações, e o viés da difusão, pela inclinação média dos usuários alcançados a partir de usuários de cada inclinação, numa simulação com modelo epidêmico (SIR, retomado no Capítulo 11). Contrastam plataformas em que o usuário pode ajustar o feed (Reddit) com as que não oferecem essa opção (Facebook e Twitter) e atribuem à baixa moderação do Gab o que descrevem como discurso de ódio generalizado. No Facebook (vacinas) e no Twitter (aborto), a inclinação do usuário se correlacionou fortemente com a de seus vizinhos e a informação alcançou sobretudo quem tinha inclinação parecida; no Reddit e no Gab, os usuários formaram uma única comunidade, inclinada à esquerda (Reddit) ou à direita (Gab). Como as diferenças entre plataformas podiam refletir os temas, os autores compararam Facebook e Reddit num tema comum, o consumo de notícias, e a segregação foi maior no Facebook. A Tabela 6.1 resume o material e os resultados dos três estudos que se seguem.

Nyhan et al. (2023) medem outra coisa: o que chega ao feed de toda a população adulta ativa do Facebook nos Estados Unidos. Só 20,6% dos usuários recebem mais de 75% do conteúdo de fontes de mesma inclinação, e só 32,2% têm um quarto ou mais vindo de fontes cruzadas; segundo os autores, os resultados não são consistentes com os piores temores sobre câmaras de eco. No experimento, a exposição a fontes cruzadas subiu de 20,7% para 27,9%. Quem teve a exposição reduzida engajou-se mais, por exposição, com o conteúdo de

mesma inclinação que ainda via, e os autores dizem que as mudanças testadas no algoritmo não anulam a propensão dos usuários a buscar e engajar com informação congenial. Para o efeito nulo citam, entre outras explicações, que o conteúdo político congenial é fração do que se vê, que o Facebook é fração da dieta de informação e que o desenho não capta efeitos de uso prévio nem acumulados ao longo de anos. Observam também que reduzir a exposição poderia levar as pessoas a buscar essa informação em outro lugar da internet.

Bail et al. (2018), por sua vez, fizeram o experimento de sentido inverso: aumentaram a exposição a opiniões opostas. Uma empresa de pesquisa recrutou republicanos e democratas autodeclarados que visitavam o Twitter ao menos três vezes por semana; 1.652 responderam a um questionário em meados de outubro de 2017 (901 democratas e 751 republicanos), com 10 afirmações sobre políticas públicas sobre as quais deveriam concordar ou discordar numa escala de sete pontos, que mede a posição liberal-conservadora. Uma semana depois, os pesquisadores sortearam quem receberia a oferta de US\$ 11 para seguir por um mês um bot, uma conta automática do Twitter, que retuitaria 24 mensagens por dia, sem que o participante soubesse o conteúdo. Os bots, um liberal e um conservador, retuitavam mensagens sorteadas de uma lista de 4.176 contas políticas (autoridades eleitas, líderes de opinião, veículos de mídia e organizações sem fins lucrativos). Aceitaram a oferta 64,9% dos democratas e 57,2% dos republicanos, e o mesmo questionário foi aplicado de novo cerca de um mês e meio depois do primeiro. Os republicanos que seguiram um bot liberal expressaram opiniões mais conservadoras: o aumento foi de 0,12 ponto na escala pela estimativa mais cautelosa, que considera todos os sorteados para o tratamento, tenham seguido o bot ou não, e de 0,60 entre os totalmente aderentes, os que responderam corretamente a todas as perguntas semanais sobre o conteúdo dos tuítes do bot. Entre os democratas, os efeitos foram pequenos, na direção liberal, e não significativos. Os autores advertem que não identificaram o mecanismo, que os participantes foram pagos para ler mensagens de quem pensa diferente e que os achados não devem ser lidos como evidência de que a exposição a visões opostas aumenta a polarização em todos os contextos. Sua segunda hipótese pré-registrada é que pessoas expostas a mensagens que conflitam com suas atitudes tendem a contra-argumentá-las com raciocínio motivado, o que acentua diferenças percebidas e o compromisso com crenças prévias.

TABELA 6.1

Três estudos sobre câmaras de eco: o que mediram, o que observaram e o que não permitem dizer

Estudo	Material	O que mediram	O que observaram	Ressalva
Cinelli et al. (2021)	Mais de 100 milhões de conteúdos de Gab, Facebook, Reddit e Twitter; inclinação inferida de curtidas, links e comentários	Homofilia nas interações sobre um tema e viés da difusão a partir de pares de mesma inclinação	A agregação em grupos homofílicos domina a dinâmica; mais segregação no Facebook (vacinas) e no Twitter (aborto) do que no Reddit e no Gab, onde os usuários formam uma única comunidade	Mede interação e difusão, não o efeito sobre o que o usuário diz ou declara (acréscimo); a inclinação é inferida
Nyhan et al. (2023)	Dados de 2020 de toda a população adulta ativa do Facebook nos EUA; experimento de três meses com 23.377 usuários	Parcela de conteúdo de fontes de mesma inclinação e de fontes cruzadas no feed; oito medidas de atitude pré-registradas	Mediana de 50,4% (mesma inclinação) e 14,7% (cruzadas); exposição a fontes de mesma inclinação de 53,7% para 36,2%, sem efeito mensurável nas atitudes	Conteúdo político congénial é fração do que se vê; o Facebook é fração da dieta; o desenho não capta efeitos de uso prévio nem acumulados ao longo de anos
Bail et al. (2018)	1.652 respondentes (901 democratas, 751 republicanos) do Twitter; US\$ 11 para seguir por um mês um bot que retuitava 24 mensagens por dia	Opiniões políticas em escala de sete pontos depois do mês	Republicanos tratados: aumento de 0,12 ponto (estimativa mais cautelosa) e 0,60 (totalmente aderentes), na direção conservadora; democratas: efeitos pequenos, na direção liberal, não significativos	Mecanismo não identificado; participantes pagos para ler mensagens de quem pensa diferente; não vale para todos os contextos

Nota. Os números vêm dos artigos. Cinelli et al. (2021) e Bail et al. (2018) foram lidos em XML do PMC, sem paginação; as páginas de Nyhan et al. (2023) são as de Nature. “(acréscimo)” marca ressalva que não é dos autores.

Os três trabalhos parecem divergir, e em parte divergem, mas medem coisas diferentes: com quem se interage sobre um tema (Cinelli et al., 2022), o que chega ao feed inteiro, em todos os assuntos (Nyhan et al., 2023) e o que declara quem é forçado a ver o outro lado (Bail et al., 2018). É possível, sem contradição, que um usuário interaja de modo homofílico sobre um tema, receba um feed pouco homogêneo no conjunto e se desloque, ao ver o outro lado, para onde já estava. Esta é interpretação, e não achado de nenhum dos três.

O que têm em comum importa mais: nenhum mede as consequências que a audiência arranja para o que o falante diz. Exposição, interação e engajamento dizem o que chega e com quem se fala, e não quem reforça o quê. O dado mais próximo é o de Nyhan et al. sobre engajamento: quem viu menos conteúdo congenial engajou-se mais, por exposição, com o que ainda via. Cabe a leitura de que a propensão a buscar o congenial é função de uma história de consequências sociais e não do que o feed oferece, e cabe também a hipótese rival, de que o usuário simplesmente gosta do que gosta. O estudo não distingue as duas.

■ DADO	Há homofilia forte nas interações sobre temas específicos no Facebook e no Twitter (Cinelli et al., 2021); o feed do usuário mediano do Facebook tem metade de fontes de mesma inclinação, só um quinto dos usuários passa de 75% e, reduzida a exposição congenial em um terço, não houve efeito mensurável nas atitudes (Nyhan et al., 2023); a exposição forçada a mensagens do campo oposto deslocou republicanos na direção conservadora e não deslocou democratas de modo significativo (Bail et al., 2018).
■ INTERPRETAÇÃO	Os três estudos medem o que chega, com quem se interage e, em dois deles, o que se declara depois de mudar a exposição. Não medem as consequências que a audiência arranja para o que o falante diz nem o que da correção chega a ele, de modo que não decidem se a bolha opera como comunidade verbal.
■ HIPÓTESE	Se a bolha opera como subcomunidade verbal, o efeito sobre o que se diz depende de quem reforça e de quanto do que vem de fora chega ao falante, e não só do que ele vê. Nenhum estudo do capítulo testa isso.

6.3 “Realidades virtuais”: quando o controle verbal substitui o contato com o evento

Esta seção não tem dado: é a análise teórica de Guerin (1992) e a extrapolação feita aqui. Guerin descreve três vias pelas quais o conhecimento socialmente construído se desliga do ambiente. Na primeira, o controle social do tato passa a um grupo que não reforça o relato correto do ambiente. Na segunda, intraverbais de escopo específico são reforçados pelo grupo como se fossem fatos lembrados. Na terceira, as consequências funcionais generalizadas ficam generalizadas demais e mantêm o tato de modo indiscriminado. Guerin ilustra a segunda via com enunciados como “reatores nucleares são inseguros”: a maioria das pessoas não tem experiência direta com esses eventos, e os enunciados se mantêm por consequências sociais. Assim, em condições sócio-econômicas que favoreçam o uso de combustíveis fósseis (contexto político favorável, preços internacionais baixos, conflitos geopolíticos etc.), haverá maior propensão de reforço para essas afirmações a despeito de qualquer informação factual adicional sobre a segurança dos reatores nucleares. Em 1994 ele retoma o exemplo ao tratar de atitudes sobre eventos que a pessoa nunca experienciou, e que por isso são intraverbais.

Um ponto de Guerin merece ser dito com todas as letras, porque afasta uma leitura errada: mesmo quando o que se diz é contrafactual, as contingências envolvidas são reais, embora sociais (1992). Quem afirma que há um elefante cor-de-rosa na sala está sob controle de contingências sociais, e não do elefante. A expressão “realidades virtuais”, do sumário deste livro, é um atalho para essas vias, e nenhuma das passagens verificadas de Guerin a usa; “virtual” não quer dizer, aqui, digital, e o artigo é anterior às redes sociais.

Duas passagens importam para o ambiente digital. Na primeira, Guerin escreve que meios de comunicação de massa podem sustentar representações sociais contrafactuais mesmo sem um grupo social, e que há potencial para comunidades verbais reforçadoras existirem só na mídia, e não presencialmente. Na segunda, define o primeiro limite do conhecimento socialmente construído: a extensão em que o comportamento é mantido por controle de estímulos não social. A consequência que se tira do primeiro limite é uma extrapolação: onde o falante tem contato direto com o evento, o relato fica sob controle dele, e a distância entre o que o grupo reforça e o que o evento mostra tende a se tornar visível. Guerin discute os limites também pela distinção entre comportamento governado por regras e por contingências (1992); o Capítulo 3 trata dela. O segundo limite é o assunto da seção seguinte.

6.4 Contracontrole e esquiva de contracontrole

Guerin (1992) aponta um limite à extensão em que o comportamento pode ser mantido por uma comunidade verbal: forte pressão de conformidade produz, em geral, contracontrole quando especifica contingências contrafactuais, a menos que se restrinja a um grupo pequeno e coeso, o que só funciona quando os benefícios de pertencer ao grupo superam os de relatar com exatidão. A passagem pede um termo e uma pergunta. O termo é contracontrole. A pergunta é o que um grupo pode fazer para que as condições do contracontrole não se formem.

CONCEITO EM USO

Contracontrole (Delprato, 2002)

Delprato (2002) define contracontrole como episódios comportamentais compostos de condições controladoras aversivas socialmente mediadas e de respostas de fuga ou esquiva que não reforçam o comportamento controlador. Não é um princípio básico do comportamento: é sempre fuga ou esquiva, e só é singular porque o sujeito, diante de estimulação social ou interpessoal aversiva, responde para se opor ao controle em vez de reforçá-lo cedendo. Entre as formas estão sair do alcance do controlador, atacar e resistir passivamente e, entre muitas outras, o absenteísmo, a crítica, a sabotagem, o não cumprimento de pedidos, a desatenção e o abandono. Ocorre em dois níveis: fuga ou esquiva de consequências específicas de curto prazo, ou de contingências coercitivas generalizadas e duradouras. Delprato associa o contracontrole eficaz a condições do controle: ele deve ser conspícuo e equilibrado com o contracontrole, as fontes devem ser identificadas e o comportamento delas, acessível aos controlados; o controle disfarçado e a delegação do controle devem ser evitados, porque tornam o comportamento do controlador inacessível.

Dois pontos do texto importam para o que segue. O primeiro é que o contracontrole é resposta a controle aversivo. Delprato (2002) escreve que as pessoas valorizam a liberdade porque ela se refere a métodos de controle não aversivos, que por isso não ocasionam contracontrole, e que quem tem o comportamento sob reforçamento positivo diz sentir-se livre porque as variáveis controladoras podem ser inconspícuas. Para a bolha, se a hipótese da seção 6.1 estiver certa, a consequência é que o controle exercido sobre o membro, feito em boa parte de aprovação, tende a não ocasionar contracontrole dentro do grupo, como na pequena comunidade coesa de Guerin. O contracontrole que interessa vem de fora ou de dissidentes: a correção de uma agência, a crítica

de quem discorda, a moderação de uma plataforma, a saída de um membro. O segundo ponto é a lista de condições citadas pelo autor e examinadas seguir, que o capítulo usa como régua.

Há ainda uma questão de sentido. Em Delprato, o contracontrole é a resposta de quem é controlado. A esquiva de contracontrole, como este capítulo a usa, é o outro lado da relação: o que o falante, ou a comunidade que o reforça, faz que reduza a chance de que essa resposta ocorra ou sua eficácia. É uma extensão proposta aqui, e o quadro abaixo diz isso sem rodeios.

CONCEITO EM USO

Esquiva de contracontrole (proposta aqui, a partir de Delprato, 2002)

Repertório do falante, ou da comunidade que o reforça, mantido por reduzir a probabilidade ou a eficácia do contracontrole: a correção, a crítica, a denúncia, a moderação, a saída. O termo não é de Delprato nem de outra fonte lida. Chamar um repertório de esquiva exige mostrar a contingência: que a correção é aversiva ao falante ou à comunidade, que o repertório a precede ou a evita e que é mantido por isso, e não apenas pela aprovação da audiência (seção 6.8). Quatro repertórios são tratados: forma, fonte, lugar e pessoas.

Forma. Guerin (1994) escreve que, havendo consequências negativas para certas declarações de atitude, elas podem ser evitadas de vários modos por meio de autoclíticos. Relatar atitudes é reforçado quando se apresentam crenças de apoio, mas só se a comunidade verbal concorda com elas; diante de audiência que não crê nelas, a crença seria omitida. O Capítulo 2 (seção 2.8) argumentou que um enunciado que passa da forma factual para a opinativa deixa de ser checável, e Guerin mostra que enunciados de atitude podem funcionar como mandos. A leitura como esquiva é que a forma tira o controle da conspicuidade: o enunciado que age sobre o ouvinte sai como opinião. O dado de Guerin, porém, tem a audiência como variável, o que sugere que a proteção da forma é mais necessária nas franjas do que dentro da bolha. Para um adepto inveterado de seu grupo, o falante diz “vacinas não funcionam”, enquanto que para pessoas na fronteira de seu grupo, ele afirma “não acredito que as varinhas funcionem”.

Fonte. Evangelista e Bruno (2019) fizeram observação participante em ao menos três grupos políticos de WhatsApp e dois de Telegram, até o dia da eleição de 2018 e nas semanas anteriores. Nos grupos claramente identificados com Bolsonaro, observaram que se repetia que não se devia acreditar na mídia tradicional, controlada por comunistas e esquerdistas, e que só alguns youtubers, jornalistas, blogueiros e políticos mereciam confiança. Depois da eleição, o conteúdo da mídia tradicional crítico ao novo governo tornou-se menos

comum nos grupos, mesmo quando vinha de comentaristas considerados de direita, e antes o material crítico ao PT era compartilhado. Para os autores, a estratégia parece ser legitimar os grupos como fontes mais confiáveis que a mídia tradicional. Na pandemia, Bonow Soares et al. (2021), que analisaram as mensagens mais compartilhadas de mais de 500 grupos políticos públicos de WhatsApp em março e abril de 2020 (802 mensagens com desinformação sobre a Covid-19), dizem que o argumento político foi usado para enquadrar como criados pela oposição todos os discursos que desacreditavam Bolsonaro. Ricard e Medeiros (2020) registram, segundo a comissão parlamentar mista de inquérito (CPMI) que desde 2019 investiga a desinformação no Congresso Nacional, que parte da desinformação atacava governadores, mídia e especialistas em saúde para deslegitimar quem apoiava o isolamento. A mudança dentro do mesmo tipo de fonte, compartilhada ou não conforme a quem critica, é o dado mais próximo de controle por audiência: o que se compartilha varia com o alvo da crítica, e não com a história da fonte.

Lugar. No relatório de 2026 da Agência Lupa, as verificações eleitorais são as de conteúdos virais ou potencialmente virais, sem autoria identificada, nos ciclos eleitorais de 2018 a 2024 (286 no total). O Facebook aparecia em 76,5% das verificações de 2018, o WhatsApp em 92,4% das de 2022 e em 100% das verificações eleitorais analisadas em 2024 (Agência Lupa, 2026). A agência lê o dado como migração dos conteúdos enganosos, à medida que as redes abertas ampliaram a moderação e passaram a sofrer maior escrutínio público, para aplicativos de mensagem e ambientes semiprivados, onde a autoria é mais difícil de identificar e o monitoramento enfrenta limitações maiores. Recuero et al. (2025), que analisaram 811 checagens de três agências brasileiras sobre as eleições de 2018 e 2022, leem a mudança do Facebook para o WhatsApp e o Twitter, entre 2018 e 2022, como adaptação a espaços mais fechados, em que é mais difícil circularem contradiscursos, e Bonow Soares et al. (2021) escrevem que o WhatsApp é canal-chave de difusão de informação falsa porque não pode ser rastreado nem desmentido publicamente por checadores. Os denominadores são pequenos (51 verificações em 2018, 131 em 2022 e 54 em 2024, lidos no gráfico), e a própria Lupa avisa que mantém uma tipline no WhatsApp e integra programas de verificação da Meta, de modo que os resultados sobre plataformas são um retrato do universo que ela monitora.

Pessoas. Nos grupos que Evangelista e Bruno (2019) observaram, comentários críticos levavam à remoção do autor e à acusação de ser petista infiltrado; telefones de acusados circulavam e os moderadores eram avisados para excluí-los. Havia pouca discussão, e poucos membros respondiam por quase todo o conteúdo enviado ou encaminhado. Os autores apontam a invisibilidade dos atores que produzem, monitoram, distribuem ou dirigem os conteúdos: a mai-

oria dos usuários desconhece a autoria do que recebe e não pode verificá-la. A Lupa escreve que a ausência de autoria identificável dos ataques a candidatos dificulta a responsabilização e favorece a rápida disseminação desses conteúdos (Agência Lupa, 2026). Os números sobre a concentração da origem das imagens são de Resende et al. (2019, citados por Evangelista e Bruno, 2019). Esses pesquisadores monitoraram grupos públicos de WhatsApp escolhidos pela correspondência entre o endereço de convite e um dicionário de nomes de políticos, de partidos e de palavras associadas a extremismo político: 141 grupos durante a greve dos caminhoneiros e 364 durante a eleição. Um método automático identificou as imagens compartilhadas que já haviam sido recusadas por serviços de checagem, e a elas se somaram 15 que a Lupa tinha identificado como desinformação. Essas 85 imagens com desinformação foram compartilhadas oito vezes mais do que outras 69.590, verdadeiras ou não denunciadas a nenhuma agência independente; eram 1% das imagens e apareceram em 44% dos grupos monitorados na campanha, e um pequeno número de grupos parecia originar grande parte delas.

TABELA 6.2

Quatro repertórios de esquiva de contracontrole: a condição de Delprato que cada um enfraquece e as fontes que o descrevem

Repertório	Condição de contracontrole eficaz que enfraquece (Delprato, 2002)	Fontes que o descrevem
1 Forma	Controle conspícuo: o enunciado que age sobre o ouvinte sai como opinião ou atitude	Guerin (1994)
2 Fonte	Nenhuma: não consta da lista de Delprato; acréscimo aqui proposto	Evangelista e Bruno (2019); Bonow Soares et al. (2021); Ricard e Medeiros (2020)
3 Lugar	Comportamento do controlador acessível aos controlados	Agência Lupa (2026); Recuero et al. (2025); Bonow Soares et al. (2021)
4 Pessoas	Fontes identificadas; sem delegação do controle	Evangelista e Bruno (2019); Agência Lupa (2026)

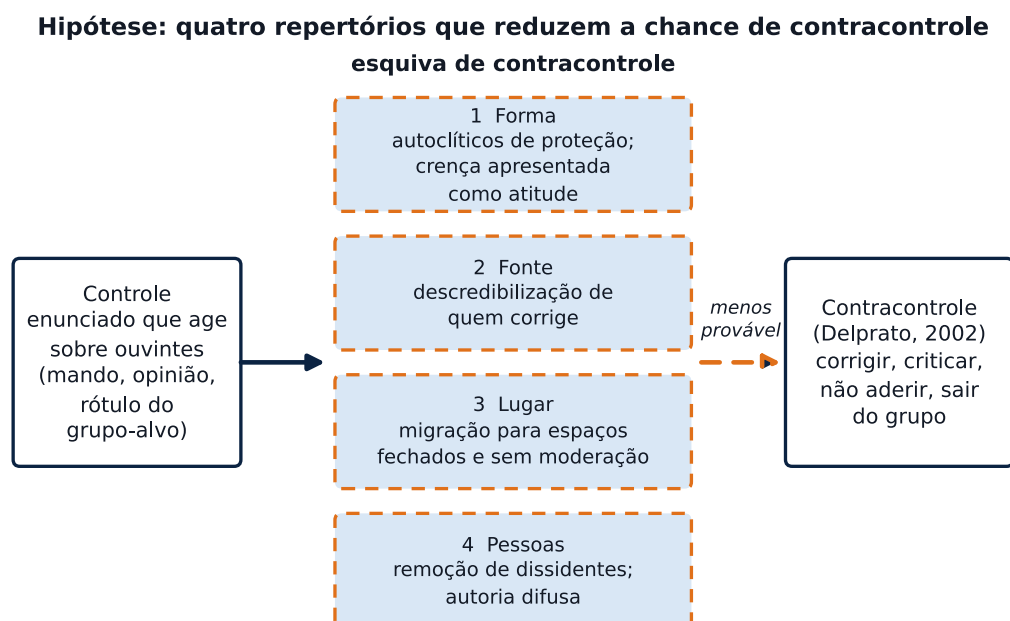
Nota. Páginas do PDF dos periódicos e do relatório da Lupa. A classificação como repertório de esquiva é interpretação aqui proposta; as ressalvas de cada fonte estão no texto da seção.

A figura resume a hipótese e não a prova. Os repertórios 1, 3 e 4 enfraquecem condições da lista de Delprato: a forma torna o controle menos conspícuo, o lugar torna o comportamento do controlador menos acessível e a autoria difusa torna a fonte menos identificável, sendo a remoção de dissidentes a eliminação de quem faria o contracontrole. O repertório 2 não está na lista; age sobre a eficácia do contracontrole e não sobre as condições dele, desqualifi-

cando quem o emite, e é por isso que é tratado como acréscimo. Que os quatro funcionem como esQUIVA no sentido técnico, mantidos por evitar a correção, é a parte que nenhuma fonte do capítulo testa: elas descrevem os repertórios e, quando muito, a leitura dos próprios autores, sem registrar a correção e o que o falante ou o grupo fez depois.

FIGURA 6.2

Controle, contracontrole e os repertórios de esQUIVA: uma hipótese



Os repertórios 1, 3 e 4 enfraquecem condições que Delprato (2002, p. 196) associa ao contracontrole eficaz (controle conspícuo, comportamento do controlador acessível, fonte identificada). O repertório 2 é acréscimo aqui proposto.

Tracejado: hipótese aqui proposta. O termo "esQUIVA de contracontrole" não é de Delprato.

Nota. Esquema original do autor, sem dados. Tracejado: hipótese. O termo "esQUIVA de contracontrole" não é de Delprato (2002). Os repertórios 1, 3 e 4 enfraquecem condições que ele associa ao contracontrole eficaz; o repertório 2 é acréscimo aqui proposto.

■ DADO	Em grupos políticos de WhatsApp e Telegram observados em 2018, a crítica levava à remoção do autor, poucos membros enviavam quase todo o conteúdo, e a mídia tradicional era descrita como não confiável (Evangelista & Bruno, 2019). A parcela de verificações da Lupa com WhatsApp subiu de 2018 a 2024 (Agência Lupa, 2026). Guerin (1994) descreve que autoclínicos evitam consequências negativas de declarações de atitude.
■ INTERPRETAÇÃO	Os repertórios descritos são compatíveis com o que Delprato (2002) associa ao contracontrole ineficaz: controle disfarçado, fontes não identificadas e comportamento do controlador inacessível. Que eles reduzam a chance ou a eficácia da correção é leitura; que sejam mantidos só pela aprovação da audiência é leitura igualmente compatível.
■ HIPÓTESE	Os repertórios de forma, fonte, lugar e pessoas são mantidos, ao menos em parte, por evitarem a correção. Isso prevê que a forma da fala mude depois de checagens, que a descredibilização acompanhe ou preceda a correção e que a remoção de dissidentes ou o deslocamento de lugar a suceda. Nenhuma fonte do capítulo registra a correção e o que veio depois. Embora estes repertórios estejam sendo apresentados aqui no contexto de desinformação, é da posição dos autores que eles estão presentes em quaisquer interações sociais.

6.5 O caso brasileiro: WhatsApp, eleições e pandemia

O material brasileiro do capítulo é pouco e vem de desenhos diferentes. A Tabela 6.3 o reúne antes de qualquer leitura.

TABELA 6.3

Material brasileiro do capítulo: o que cada fonte tem, o que registra e o que não permite dizer

Fonte	Material	O que registra	O que não permite dizer
Evangelista e Bruno (2019)	Observação participante em ao menos três grupos políticos de WhatsApp e dois de Telegram, até a eleição de 2018 e semanas antes; dados de Resende et al., citados pelos autores	Remoção de quem criticava; poucos membros enviavam quase todo o conteúdo; mídia tradicional descrita como controlada por comunistas e esquerdistas; autoria invisível	Poucos grupos, sem amostragem descrita: não mede prevalência; os autores dizem não poder confirmar plenamente o uso eficaz de microsegmentação

Fonte	Material	O que registra	O que não permite dizer
Bonow Soares et al. (2021)	Mensagens mais compartilhadas de mais de 500 grupos políticos públicos de WhatsApp (mar-abr/2020): 802 com desinformação sobre Covid-19, segundo os autores	41% teorias da conspiração, 39% distorções, 20% informação fabricada; entre os tópicos (não excludentes), 36% governadores e prefeitos com interesses ocultos contra Bolsonaro e 21% a mídia	Só grupos públicos, só texto, cerca de um terço das mensagens compartilhadas ao menos 20 vezes por dia; muitos grupos são de apoiadores de Bolsonaro
Ricard e Medeiros (2020)	Comentário, sem dados próprios; depoimentos CPMI	Parte da desinformação atacava governadores, mídia e especialistas em saúde para deslegitimar quem apoiava o isolamento	Não é estudo empírico; a comissão ainda não havia divulgado evidência de coordenação direta pelo chamado Gabinete do Ódio ou por membros do governo
Agência Lupa (2026)	286 verificações eleitorais (conteúdo viral sem autoria identificada), de 2018 a 2024	Facebook em 76,5% (2018); WhatsApp em 92,4% (2022) e 100% (2024); 73,1% sobre urnas, apuração, TSE ou resultados; 50% envolvendo instituições públicas	Retrato do universo que a Lupa monitora (tipline no WhatsApp, programas da Meta); categorias não excludentes
Recuero et al. (2025)	811 checagens de três agências, eleições de 2018 e 2022	Mudança de plataforma lida como adaptação a espaços mais fechados; ambas as campanhas aparecem como produtoras e vítimas	Estudo de checagens, não de grupos; leitura dos autores

Nota. Páginas de Evangelista e Bruno (2019), Bonow Soares et al. (2021), Ricard e Medeiros (2020), Agência Lupa (2026) e Recuero et al. (2025): as do PDF dos periódicos e do relatório. Os denominadores da Lupa, de 51 a 131 verificações por ano, são lidos no gráfico da p. 80.

Duas coisas se leem na tabela. A primeira é o que se repete entre fontes de desenhos diferentes: a desqualificação de quem está fora do grupo, de quem corrige ou de quem apoia uma política contrária (Evangelista e Bruno; Bonow Soares et al.; Ricard e Medeiros) e a autoria difícil de identificar (Evangelista e Bruno; Agência Lupa). Na pandemia, Bonow Soares et al. (2021) registram que a desinformação nos grupos monitorados esteve frequentemente ligada a teorias da conspiração, como supostas tramas de esquerdistas, do STF e da mídia contra Bolsonaro. Dois codificadores classificaram as 802 mensagens, com um terceiro para desempatar: 41% eram teorias da conspiração, 39% distorções e 20% informação fabricada, em categorias mutuamente exclusivas e com confiabilidade entre codificadores (alfa de Krippendorff, medida de concordância)

de 0,601, o que os autores chamam de moderada a substancial. Entre os tópicos, que podiam se sobrepor, 36% tratavam de governadores e prefeitos com interesses ocultos contra Bolsonaro e 21% da mídia. Na eleição, metade das verificações da Lupa envolve instituições públicas, o Judiciário, as Forças Armadas ou órgãos de segurança, frequentemente apresentados como participantes de supostas conspirações ou irregularidades (Agência Lupa, 2026).

Há uma hipótese em que esses números se encaixam, e ela é aqui proposta: se as instituições que apuram e verificam a eleição, a corte eleitoral, o Judiciário, a imprensa, são as fontes do contracontrole a alegações sobre eventos eleitorais, apresentá-las de antemão como participantes de conspirações é um modo de enfraquecer o contracontrole que viria delas, o repertório de fonte da seção 6.4. A leitura rival é que os conteúdos apenas tratam do que há para tratar numa eleição: o assunto das urnas é o assunto da eleição, e as instituições são seu objeto. O material não separa as duas leituras.

A segunda é o equilíbrio. Quase todo o material brasileiro do capítulo trata de grupos e campanhas de um campo político, o de apoiadores de Bolsonaro, e isso decorre do desenho das fontes (Bonow Soares et al., 2021; Evangelista e Bruno, 2019), e não de uma escolha de campo, já que o mesmo método se aplica a todos. Recuero et al. (2025) registram que a literatura que citam aponta mais desinformação produzida por apoiadores de Bolsonaro nas duas eleições, mas que ambas as campanhas aparecem como produtoras e vítimas de desinformação. Evangelista e Bruno (2019) dizem que, na amostra de Resende et al., não há evidência de que a coalizão de esquerda tenha usado as mesmas táticas, embora haja relatos de que a campanha de Haddad também usou o WhatsApp. Os dois registros não decidem se o fenômeno é assimétrico: dizem que a pergunta, para os repertórios de esquiva, não foi feita com amostra comparável. A hipótese do capítulo é de mecanismo: os repertórios são definidos por sua relação com a correção, e não por partido, e deveriam aparecer onde houver um grupo coeso sob controle de uma audiência que o reforça. Ausência de dado não é ausência de fenômeno, e o capítulo não afirma simetria; afirma que o dado não permite dizer.

■ DADO	Em grupos políticos de mensagens observados em 2018 e em grupos públicos de WhatsApp monitorados em 2020, descrevem-se remoção de críticos, poucos produtores, autoria invisível, desconfiança da mídia tradicional e, na pandemia, teorias da conspiração envolvendo governadores, mídia, esquerdistas e STF (Evangelista & Bruno, 2019; Bonow Soares et al., 2021). Metade das verificações eleitorais da Lupa envolve instituições públicas, frequentemente apresentadas como participantes de supostas conspirações (Agência Lupa, 2026).
■ INTERPRETAÇÃO	O que se descreve é compatível com os quatro repertórios da seção 6.4, e a classificação é aqui proposta. As fontes tratam, em sua maioria, de um campo político, por desenho.
■ HIPÓTESE	Os repertórios aparecerão, com a mesma função, em grupos de outros campos quando a mesma regra de amostra for aplicada a eles (ver o protocolo do Capítulo 2, seção 2.9). Não há dado sobre isso no capítulo.

6.6 Quem curte, quem responde: uma análise exploratória de dois vídeos políticos no Instagram

Em outubro de 2026, na campanha do segundo turno da eleição presidencial, foi publicada no perfil de Instagram de um dos autores (@marcioborgesmoreira) uma série de vídeos curtos sobre política. O perfil tinha 11.550 seguidores e vivia de carrosséis de divulgação científica em Análise do Comportamento, vistos em geral por algumas centenas de contas. Dois vídeos da série saíram da curva: “Eram só vovós com Bíblia?”, sobre o 8 de janeiro de 2023, teve 1,29 milhão de visualizações, e “Quanto custava o litro da gasolina em cada ano?”, 735 mil. O autor esperava ataques. O tema, a polarização da campanha e um público cem vezes maior que a lista de seguidores pareciam reunir as condições de uma audiência negativa, no sentido de Skinner retomado na seção 10.1. Os ataques não vieram: no Direct, nada; nos comentários, pelo que se via, quase só aprovação. A impressão foi a de que os vídeos tinham ido para um milhão de pessoas da bolha do autor, a do campo de Lula.

Esta seção examina essa impressão com os dados do próprio perfil, e o caso interessa ao capítulo por uma razão precisa. Os estudos da seção 6.2 medem o que chega ao feed e com quem se interage; nenhum mede as consequências que a audiência arranja para o que se diz nem o que acontece com quem discorda. Num perfil, quem publica tem acesso a parte disso: quantas contas viram, quantas eram seguidoras, o que cada pessoa escreveu, quantas curtidas

cada comentário recebeu e que respostas provocou. O caso não testa a hipótese do capítulo, porque é um perfil e um episódio, mas mostra o que se ganha ao descrever a bolha pelas consequências e onde ainda faltam dados.

Há uma ressalva de origem, e ela vem antes dos números: os vídeos são de um dos autores, que tinha posição na campanha. Isso não muda as regras de leitura do livro (dado, interpretação e hipótese separados; o mesmo critério para todos os campos), mas o leitor deve considerar ao julgar a classificação dos comentários, que é interpretação.

O material e como foi obtido

O material cobre os 74 posts publicados no perfil entre 21 de setembro e 21 de outubro de 2026 e foi lido em 21 de outubro, entre 20h e 22h (horário de Brasília). Veio de três fontes. A Graph API do Instagram, a interface que a Meta oferece às contas profissionais, deu a data, o tipo, a legenda, as curtidas e a contagem de comentários de cada post e, nos dois vídeos, o texto de todos os comentários e das respostas a eles. A chave de acesso disponível não tinha permissão para ler métricas; por isso visualizações, alcance, a divisão entre seguidores e não seguidores, compartilhamentos, salvamentos, tempo médio assistido e o público por idade e gênero foram lidos nas páginas de cada post do Meta Business Suite, o painel da própria Meta, e o perfil dos seguidores, na página de público do mesmo painel.

Duas medidas da plataforma precisam ser explicadas antes de usadas. Alcance é o número de contas diferentes que viram o post; visualização é cada vez que ele começa a ser exibido, e por isso as visualizações passam do alcance (no vídeo do 8 de janeiro, 1,37 por conta). A plataforma também reparte visualizações e alcance entre seguidores e não seguidores do perfil, e as duas repartições não fecham: a página do vídeo do 8 de janeiro atribui 224.100 visualizações a seguidores, mais do que as duas contas que o publicaram (11.550 e 5.198 seguidores) produziram mesmo com repetições, enquanto a página da conta registra 5.202 seguidores alcançados no mês inteiro. Como a Meta não explica a diferença, a repartição é usada aqui só pela direção, que as duas medidas concordam em mostrar.

Os comentários vieram incompletos, e a falta é ela mesma um dado. Dos 2.529 que a plataforma contava no vídeo do 8 de janeiro, a API devolveu 2.076 (1.901 de primeiro nível e 175 respostas), escritos por 1.778 contas; dos 962 do vídeo da gasolina, 876 (637 e 239), de 705 contas. Faltam 453 (18%) e 86 (9%). Nenhum dos devolvidos estava marcado como oculto, e a API não diz o que falta: comentários apagados por quem os escreveu, contas restritas e comentários retidos pelos filtros automáticos da plataforma são explicações possíveis, que

o painel do aplicativo permitiria conferir e que não foram conferidas. Os comentários, com o nome de usuário, ficaram só no servidor do autor; aqui não aparece nome, e as falas hostis vêm parafraseadas.

Cada comentário recebeu três códigos, pelas regras do Quadro 6.1: a posição em relação ao vídeo, o campo político que declara e, se houver, a hostilidade e seu alvo. A leitura foi feita por um só codificador, o assistente de inteligência artificial usado na preparação deste livro, comentário por comentário (2.457 deles); outros 495, só com menção a outra conta, só com figurinha, só com emojis de aplauso ou riso ou só com o slogan de um candidato, foram classificados por regra. Não houve segunda codificação independente, e a concordância entre codificadores não foi medida; os números da Tabela 6.5 valem com essa ressalva.

QUADRO 6.1

Regras de codificação dos comentários

Posição em relação ao vídeo (uma por comentário):

- **Adesão ao conteúdo:** concorda com o que o vídeo mostra ou sustenta, elogia, agradece, acrescenta argumento ou fato na mesma direção ou ironiza a versão que o vídeo contesta.
- **Manifestação de campanha:** voto, slogan, hashtag ou pedido de voto, sem referência ao conteúdo do vídeo.
- **Contestação:** discorda, corrige, contra-argumenta, relativiza ou contrapõe outro caso (“e o MST?”), qualquer que seja o campo de quem contesta.
- **Ataque ao autor:** insulto, desqualificação pessoal ou acusação de má-fé dirigidos a quem fez o vídeo.
- **Sem posição:** só menção a outra conta, figurinha, emoji ambíguo, pergunta neutra, relato de preço local sem contestação, assunto alheio.

Campo declarado: Lula e esquerda, Bolsonaro e direita, ou nenhum; só quando o comentário o diz (voto, slogan, apoio ou ataque nominal a candidato, rótulo do outro campo, como “gado” ou “petralha”).

Hostilidade e alvo: insulto, rótulo depreciativo de pessoa ou grupo, desumanização, desejo de dano ou de punição além da que a Justiça decidiu, zombaria de atributo pessoal (idade, aparência). Alvos: participantes do 8 de janeiro; eleitores ou militantes de direita; políticos de direita; Lula, o PT, a esquerda ou seus eleitores; o autor; outro comentarista; outro. Não contam termos que descrevem atos (“vândalos”, “golpistas”, “criminosos”) nem juízos sobre ideologias (“o bolsonarismo é uma seita”). “Discurso de ódio” não é código: no livro, é rótulo de uma instância externa (Capítulo 5).

Ironia: só é reconhecida com marcador (riso, aspas, “contém ironia”) ou quando a leitura literal contradiz o próprio comentário; sem isso, vale a leitura literal.

Grafia disfarçada: letra trocada por número, @ ou asterisco, ou palavra partida por pontos ou espaços (“c@gona”, “L U L A”). Uma regra automática acha os candidatos; a leitura decide.

O que os números do perfil mostram

O perfil tinha uma rotina antes da série, e é contra ela que os dois vídeos se medem. Nos 45 carrosséis de ciência do período, a mediana foi de 464 visualizações e 202 contas alcançadas, e só 3,7% das visualizações vieram de quem não seguia o perfil: eram posts vistos pela base. A série política começou na noite de 14 de outubro e somou 24 Reels em sete dias. Os 14 Reels de dados, todos no formato “E quem estava na Presidência” (uma série anual de fonte oficial, com a foto de cada presidente), tiveram mediana de 3,3 mil visualizações, com 86,8% delas de não seguidores; os outros 8 Reels políticos, 2,4 mil e 76,8%. Os dois vídeos que saíram da curva somaram 2,02 milhões de visualizações, 93% de tudo o que os 74 posts tiveram no mês (Tabela 6.4 e Figuras 6.3 e 6.5).

TABELA 6.4

Os dois vídeos e a rotina do perfil: medianas por tipo de post, de 21/9 a 21/10/2026

Medida	Carrosséis de ciência (45)	Reels de dados (14)	Outros Reels políticos (8)	Vídeo do 8 de janeiro	Vídeo da gasolina
Visualizações	464	3,3 mil	2,4 mil	1.293.276	735,5 mil
Contas alcançadas	202	2,6 mil	2,0 mil	947.057	536,3 mil
Visualizações de não seguidores	3,7%	86,8%	76,8%	82,5%	93,9%
Curtidas	17	61	63	77.673	17.015
Comentários	0	1	1,5	2.529	962
Compartilhamentos	1	12,5	14	23.232	14,7 mil
Salvamentos	6	9	9	8.864	2,1 mil
Tempo médio assistido	não se aplica	16 s	17 s	26 s (de 59 s)	25 s (de 49 s)
Homens entre as contas alcançadas	49,7%	79,7%	68,4%	54,0%	84,9%
Seguidores ganhos	0	1,5	3	144	735

Nota. Medianas nas três primeiras colunas. Reels publicados de 14 a 21/10/2026; a coluna dos carrosséis cobre o mês inteiro, e a parcela de não seguidores tem 44 posts, porque a página de um deles não a mostrava. Nos posts sem publicação cruzada no Facebook, a plataforma dá o alcance arredondado (“2,6 mil”). O vídeo do 8 de janeiro saiu em colaboração com outra conta, de 5.198 seguidores. Fontes: Graph API do Instagram e Meta Business Suite, lidos em 21/10/2026; tabela completa, por post, em [verificacao/caso-instagram-2026/posts_metrics.tsv](#).

A mudança de escala aparece inteira no alcance diário da conta (Figura 6.4). Até 13 de outubro, o perfil alcançava de 200 a 800 contas por dia; o único salto do período foi um anúncio pago de cinco dias, de 30 de setembro a 4 de outubro, para um post de antes da série (R\$ 139,92, 23.502 visualizações). Em 15 de outubro, primeiro dia inteiro do vídeo do 8 de janeiro, foram 563.637 contas. Nenhum dos dois vídeos foi impulsionado. O do 8 de janeiro passou de 500 mil visualizações cerca de 19 horas depois de publicado e de 1 milhão em dois dias e meio; o da gasolina chegou a 500 mil em cerca de 39 horas e a 700 mil em quatro dias. Nos 28 dias de 23 de setembro a 20 de outubro, a conta alcançou 1,5 milhão de contas, das quais 5.202 eram seguidoras (0,3%) e 1.512.615 não eram (99,7%); 97% das interações vieram de não seguidores.

FIGURA 6.3

Os dois vídeos que saíram da curva e a série de que faziam parte

“Eram só vovós com Bíblia?”
14/10, 23h29, em colaboração



“Quanto custava o litro da gasolina?”
17/10, 11h



visualizações	curtidas
1.293.276	77.673
contas alcançadas	comentários (lidos)
947.057	2.529 (2.076)
de não seguidores	compartilhamentos
82,5% das visual.	23.232
tempo médio assistido	salvamentos
26 s de 59 s	8.864

visualizações	curtidas
735,5 mil	17.015
contas alcançadas	comentários (lidos)
536,3 mil	962 (876)
de não seguidores	compartilhamentos
93,9% das visual.	14,7 mil
tempo médio assistido	salvamentos
25 s de 49 s	2,1 mil

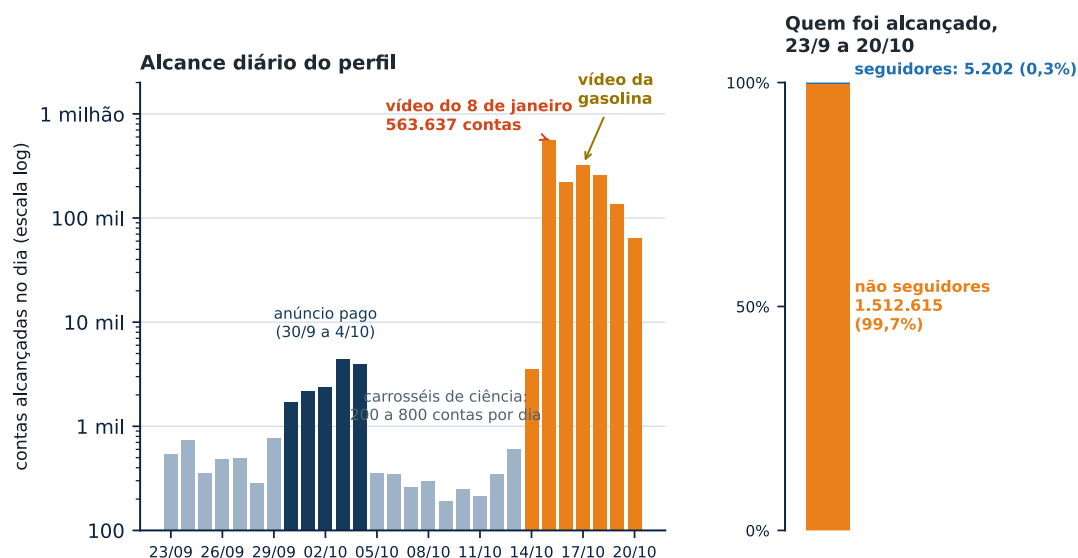
A mesma série, nos mesmos dias: oito Reels de dados e as visualizações de cada um



Nota. Capas dos posts do autor, obtidas pela Graph API. Na do 8 de janeiro, a metade de cima foi desfocada: mostra uma pessoa identificável, que não é figura pública, num quadro de vídeo com a marca de um jornal. Fotos oficiais nas demais capas: Ricardo Stuckert/PR, Roberto Stuckert Filho/PR, Beto Barata/PR e Alan Santos/PR (CC BY), creditadas nas legendas dos posts. Números do Meta Business Suite e da Graph API, lidos em 21/10/2026.

FIGURA 6.4

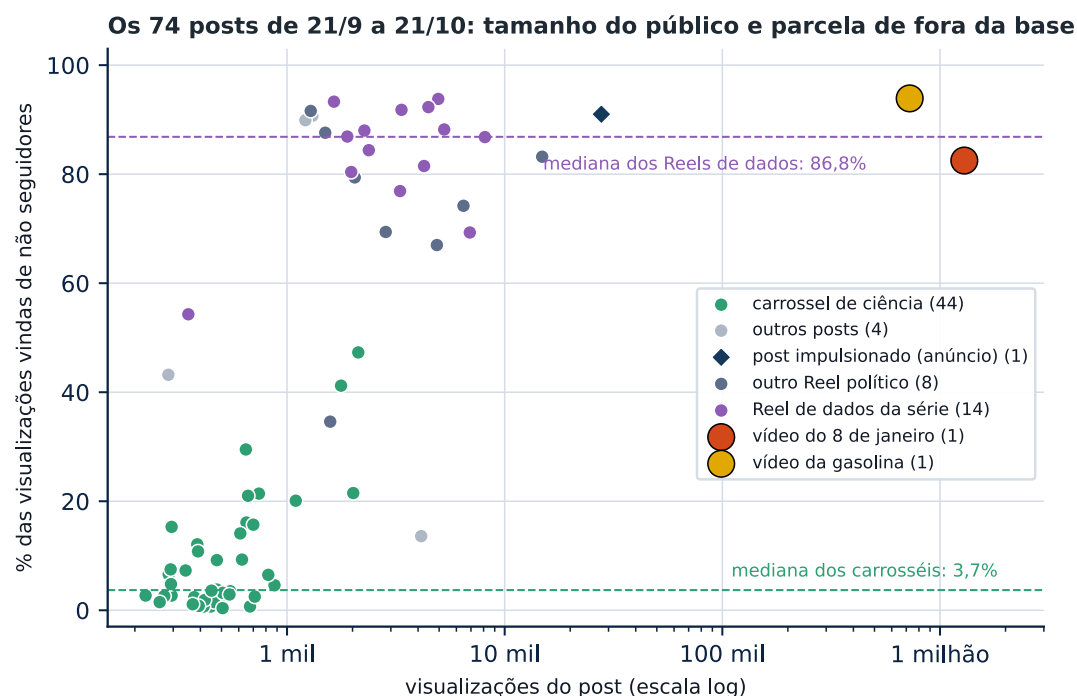
Alcance diário do perfil e parcela de seguidores e não seguidores, de 23/9 a 20/10/2026



Nota. Dias contados no fuso do Pacífico, como no painel da Meta: o vídeo do 8 de janeiro, publicado às 23h29 de 14/10 em Brasília, cai no fim do dia 14/10 desse fuso. Fonte: Meta Business Suite, visão geral da conta no Instagram, lida em 21/10/2026.

FIGURA 6.5

Os 74 posts do período: tamanho do público e parcela vinda de fora da base

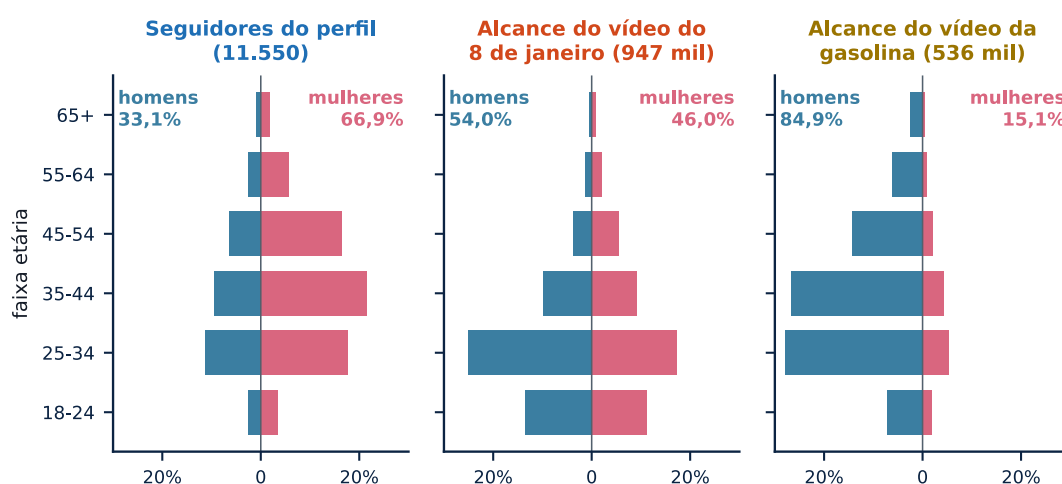


Nota. Cada ponto é um post; a parcela de não seguidores é a das visualizações, como a plataforma a reparte. "Outros posts": três Reels institucionais (dois de livros do Instituto e um da Rádio iW4) e uma imagem sobre o voto. Fonte: Meta Business Suite, páginas de cada post, lidas em 21/10/2026.

Quem foi alcançado também não era a base em outro sentido, o da composição (Figura 6.6). Os seguidores do perfil são 66,9% mulheres, e 83% deles têm de 25 a 54 anos. Entre as 947 mil contas alcançadas pelo vídeo do 8 de janeiro, 54% eram de homens, e 67% tinham de 18 a 34 anos; no vídeo da gasolina, 84,9% eram de homens. Mesmo comparado com o público que os carrosseis de ciência alcançavam, que tinha mediana de 49,7% de homens, o dos Reels de dados (79,7%) e o da gasolina eram outro público.

FIGURA 6.6

Quem segue o perfil e quem os dois vídeos alcançaram, por idade e gênero



Nota. Percentuais do total de cada público. Seguidores: página de público do Meta Business Suite; alcance: página de cada post. A plataforma infere idade e gênero do cadastro das contas. Lidos em 21/10/2026.

Duas outras medidas dizem o que essa audiência fez além de ver. Ela compartilhou e curtiu muito, mas quase não passou a seguir: o vídeo do 8 de janeiro deu ao perfil 144 seguidores novos, um para cada 6,6 mil contas alcançadas, e o da gasolina, 735; no mês, foram cerca de mil seguidores novos e 404 que deixaram de seguir. E não escreveu ao perfil: o painel registra 36 conversas iniciadas no Direct em 28 dias, 16 delas nos sete dias da série, e 12 contatos novos, nada que acompanhe um alcance de 1,5 milhão de contas. O conteúdo das mensagens não foi lido.

O que os comentários disseram

Os dois vídeos tiveram comentários de composição muito diferente (Tabela 6.5 e Figura 6.7). No do 8 de janeiro, 52,1% eram de adesão ao conteúdo, 19,3% de campanha de Lula e 26,0% sem posição; a contestação somou 2,2% (45 comentários), metade dela de quem declarava o campo de Bolsonaro, e só um comentário atacou o autor. Os comentários que declaravam o campo de Bolsonaro foram 32 (1,5%), contra 536 (25,8%) do campo de Lula. A adesão tinha for-

mas que se repetiam: ironizar a versão das “vovós” (“tadinhos”, “só tinham santos”), repetir a frase da senhora que o vídeo mostra, notar que os próprios participantes gravaram as imagens, comparar com protestos da esquerda e pedir que não houvesse anistia. A campanha vinha em números, slogans e hashtags contra o adversário; uma única conta colou a mesma mensagem de voto 30 vezes.

No vídeo da gasolina, um terço dos comentários (35,7%) contestou o vídeo, e a contestação veio de três lados. De quem declarava o campo de Lula vieram 117 comentários dizendo que o vídeo mentia porque, no governo Bolsonaro, o litro tinha chegado a 9, 10 ou 12 reais: preços de pico, de um posto e de uma cidade, contra uma média nacional anual corrigida pela inflação, que a capa e a legenda anunciavam como “em reais de hoje”. Do campo de Bolsonaro vieram 118, quase todos lembrando a pandemia, a guerra na Ucrânia e o corte de impostos de 2022, que o vídeo não mencionava. E 78, sem campo declarado, contestaram com os preços nominais da época (“em 2006 era 2,50”), muitas vezes copiando o resumo de um buscador. Os comentários do campo de Bolsonaro foram 13,9% do total, nove vezes a proporção do outro vídeo.

TABELA 6.5

Os comentários dos dois vídeos por categoria, e as curtidas que a audiência deu a cada uma

Categoria	8 de janeiro: comentários	8 de janeiro: curtidas	Gasolina: comentários	Gasolina: curtidas
Adesão ao conteúdo	1.081 (52,1%)	30.321 (91,0%)	202 (23,1%)	711 (24,3%)
Campanha de Lula	400 (19,3%)	820 (2,5%)	138 (15,8%)	414 (14,1%)
Sem posição	539 (26,0%)	2.062 (6,2%)	217 (24,8%)	1.008 (34,4%)
Contestação pelo campo de Lula	1 (0,0%)	0	117 (13,4%)	565 (19,3%)
Contestação sem campo declarado	22 (1,1%)	4 (0,0%)	78 (8,9%)	183 (6,2%)
Contestação pelo campo de Bolsonaro	22 (1,1%)	103 (0,3%)	118 (13,5%)	49 (1,7%)
Campanha de Bolsonaro	10 (0,5%)	3 (0,0%)	4 (0,5%)	0
Ataque ao autor	1 (0,0%)	0	2 (0,2%)	0
Total	2.076	33.313	876	2.930
Com hostilidade, qualquer alvo	265 (12,8%)	941 (2,8%)	110 (12,6%)	168 (5,7%)

Nota. Categorias do Quadro 6.1, com posição e campo combinados. Comentários e respostas que a Graph API devolveu (2.076 de 2.529 e 876 de 962); curtidas dadas a esses comentários até 21/10/2026, à noite. Na gasolina, a categoria “sem posição” concentra curtidas por causa de uma figurinha que recebeu 849. Codificação de um só codificador, sem medida de concordância. Fonte: `verificacao/caso-instagram-2026/comentarios_categorias.tsv`.

A hostilidade existiu nos dois vídeos, em proporção parecida (12,8% e 12,6% dos comentários), e quase toda apontou para fora. No do 8 de janeiro, os alvos foram os participantes dos atos (117 comentários: xingamentos à senhora do vídeo, zombaria da idade, pedidos de pena maior que a decidida, alguns desejos de morte), políticos de direita (75), eleitores de direita (31, com “gado” e “bolsominions”), outros comentaristas (24), Lula, o PT ou a esquerda (5) e o autor (1). No da gasolina, outros comentaristas (39, quase todos em respostas), políticos de direita (32), Lula, o PT ou a esquerda (23), eleitores de direita (13) e o autor (3). Somados os dois vídeos, quatro de 2.952 comentários (0,14%) foram hostis a quem os publicou; um deles, de quem se declarava do campo de Lula, acusava o vídeo de mentir por mostrar o preço do governo Bolsonaro mais baixo do que ele lembrava. Quem chegou com hostilidade a dirigiu, em geral, ao grupo que o vídeo criticava, e não a quem o publicou.

Quem reforça o quê: curtidas e respostas

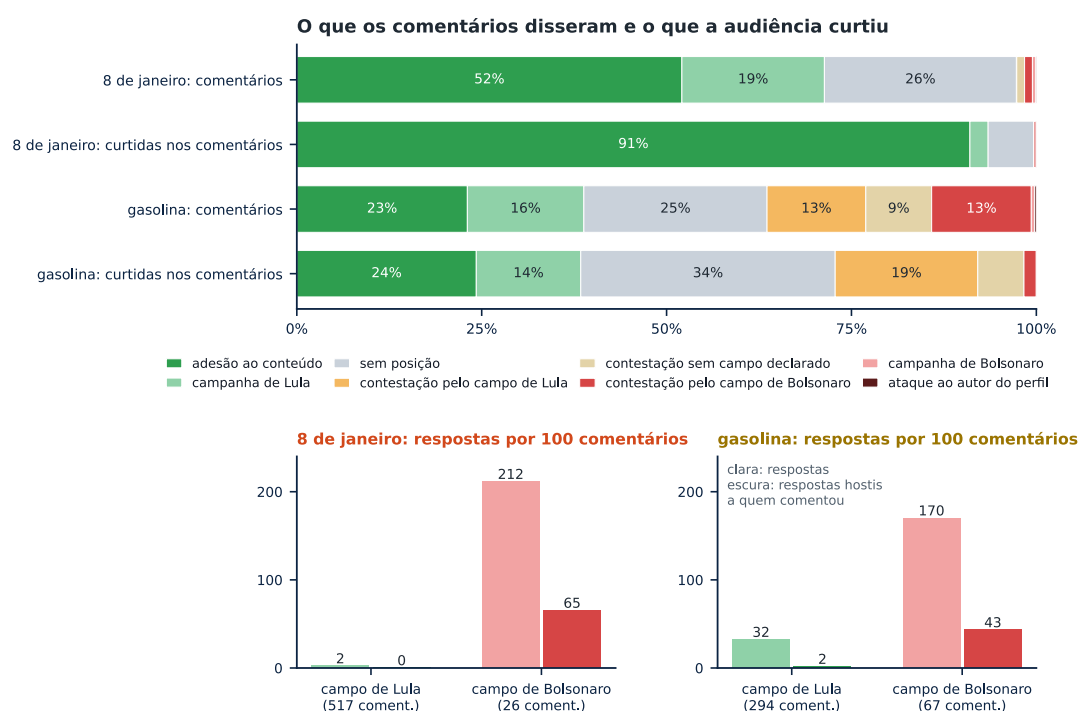
A audiência não só comenta: curte comentários e responde a eles, e essas são as consequências que ela arranja, visíveis e contadas, para o que cada pessoa escreveu. No vídeo do 8 de janeiro, os comentários de adesão receberam 91% de todas as curtidas dadas a comentários. Os quatro mais curtidos, todos de adesão, tinham 12.623, 8.608, 3.999 e 1.266 curtidas (o mais curtido dizia que “a melhor parte é que eles mesmos gravaram”); como a plataforma ordena os comentários por relevância, era isso que se via primeiro ao abri-los, e é provável que a minha impressão de “quase só aprovação” viesse dessa vitrine. No vídeo da gasolina, os dois grupos de contestação partidária tinham o mesmo tamanho, 117 e 118 comentários, mas os do campo de Lula receberam 565 curtidas e os do campo de Bolsonaro, 49. A audiência reforçou a contestação pelo campo de quem contestava, e não pela qualidade do argumento: lembrar a pandemia não é menos pertinente que lembrar um preço de pico. A correção dentro do próprio campo também foi reforçada: um comentário que dizia que no governo Dilma a gasolina nunca chegara a 4 reais teve 169 curtidas, a resposta que explicava a correção pela inflação teve 284, e quem tinha comentado respondeu “entendi”.

As respostas mostram o outro lado da mesma contingência (Figura 6.7). No vídeo do 8 de janeiro, os 517 comentários de primeiro nível do campo de Lula receberam, juntos, 12 respostas, uma delas hostil; os 26 do campo de Bolsonaro receberam 55, das quais 17 hostis a quem tinha comentado. Por cem comentários, são 2 respostas e menos de uma hostil num campo, 212 respostas e 65 hostis no outro. No vídeo da gasolina, onde a discussão foi maior para todos, a diferença persiste: 32 respostas e 2 hostis por cem comentários do campo de Lula, 170 e 43 por cem do campo de Bolsonaro.

Dois fios mostram como isso se dá. No vídeo do 8 de janeiro, uma comentarista escreveu que os participantes “aprenderam com o MST”, com dois emojis de riso. O comentário teve 94 curtidas e 47 respostas: parte argumentava (o MST produz alimentos, doou comida na pandemia), e 15 eram hostis a ela, entre elas chamá-la de ignorante e mandá-la calar-se, dizer que quem devia opinar era o marido, comentar a cor da pele dela e a profissão do marido e, numa delas, um aviso velado sobre as fotos das filhas que ela publicava. No vídeo da gasolina, um comentarista escreveu que na cidade dele o litro custava 5 reais no governo Bolsonaro e 7,80 agora. Recebeu 35 respostas, 18 delas hostis, quase todas de zombaria; a mais curtida, com 77, perguntava se havia milícia nos postos da cidade dele. Num terceiro episódio, o comentário “Flávio Bolsonaro é o nosso futuro presi..diário” (659 curtidas) foi lido por alguns como apoio a Flávio e recebeu respostas como “o seu, essa maldição”, até que alguém avisou que não tinham lido direito: a audiência respondia ao que parecia dissidência antes de lê-la inteira.

FIGURA 6.7

O que os comentários disseram, o que a audiência curtiu e o que aconteceu com quem discor- dou



Nota. Comentários e respostas devolvidos pela Graph API; curtidas até 21/10/2026, à noite. Embaixo, só os comentários de primeiro nível com campo declarado e as respostas que receberam; resposta hostil é a que tem hostilidade dirigida a quem comentou (Quadro 6.1). Codificação de um só codificador. Fonte: `verificacao/caso-instagram-2026/comentarios_categorias.tsv` e `respostas_por_campo.tsv`.

Forma, fonte e pessoas: os repertórios da seção 6.4 na caixa de comentários

Os quatro repertórios da seção 6.4 foram descritos, no material brasileiro do capítulo, em grupos de mensagens de um só campo político. Três deles aparecem na caixa de comentários destes vídeos, numa audiência alinhada com o campo oposto, e a descrição a seguir é de topografia: diz que forma tinham e que consequências se seguiram, e não que o repertório seja mantido por evitar a correção.

Forma. Vinte e oito comentários (0,9%) disfarçavam a grafia: “c@gona”, “vaga@bund0s”, “m4t4e”, “ban.did0 bom e ba.ndid0 m0.rt0”, “presi..diário”, “L U L A”. O disfarce recaía sobre xingamentos, desejos de morte e nomes de candidatos, e um comentário dizia a função às claras: evitem escrever o nome de Lula, porque o algoritmo está nos travando. É forma que reduz a chance de a moderação automática da plataforma, um controle vindo de fora da comunidade, alcançar o que se escreve, e o comentário que pedia o disfarce mostra que pelo menos uma pessoa descrevia a contingência assim. Que a moderação de fato travasse esses comentários, nenhum dado do caso mostra.

Fonte. Na gasolina, contestações dos dois campos chamaram o vídeo de mentira, e as respostas a quem discordava chamaram a discordância de fake news (“essas fake news não cola aqui”). A desqualificação da fonte do contracontrole, que a seção 6.4 descreveu nos grupos de 2018 a propósito da mídia tradicional, aparece aqui na escala do comentário, dirigida a quem traz o dado contrário.

Pessoas. Na caixa de comentários não há administrador que remova quem critica, como nos grupos observados por Evangelista e Bruno (2019), mas há o equivalente que a audiência pode arranjar: responder em massa, com zombaria e insulto, ao comentário que discorda. Os números da seção anterior são essa consequência, contada.

Lugar. Não houve deslocamento para espaço fechado; houve o movimento contrário. Comentários de campanha de Lula traziam hashtags do campo adversário (“#bolsonaro2022”, “#fláviobolsonaropresidente”), o que põe a mensagem nas buscas de quem segue essas marcas. É leitura da função, e o caso não mede se alguém do outro campo chegou ao vídeo por elas.

Era a bolha perfil?

A pergunta da abertura pede três respostas, porque “bolha” cobre três coisas diferentes. Se bolha é a lista de seguidores, os vídeos não ficaram nela: 99,7% das contas alcançadas não seguiam o perfil, e o público alcançado era mais jovem e mais masculino que a base. Se bolha é quem viu, o caso não sabe dizer:

o painel não informa a posição política de ninguém, e os 2.076 comentários lidos do vídeo do 8 de janeiro foram escritos por 0,2% das contas que ele alcançou. Se bolha é a subcomunidade verbal da seção 6.1, os ouvintes que reagem ao que se diz e arranjam consequências para isso, a resposta é sim, e ela tem números: no vídeo do 8 de janeiro, 1,5% dos comentários declaravam o campo oposto ao do vídeo, as curtidas foram para a adesão e os poucos que discordaram receberam respostas em proporção cem vezes maior que os demais, de um quarto a um terço delas hostis. Guerin (1992) escreve que mudanças no uso da linguagem podem estar sob controle de subcomunidades verbais menores que a comunidade verbal ampla; esta foi uma, montada em dois dias, que não existia como grupo antes do vídeo e não se tornou seguidora depois dele.

Como uma audiência de um milhão de contas de fora da base pode ser tão homogênea no que diz é a pergunta que o caso deixa, e ela tem quatro respostas rivais. A primeira é a entrega: o ranqueamento da plataforma decide quantos olhos e quais (seção 8.3), e, se ele usa as reações de quem já viu para escolher quem verá em seguida, a audiência vai sendo selecionada pelas próprias reações; nenhuma fonte do livro mostra o que o ranqueamento usa como entrada, e o caso também não. A segunda é a audiência negativa: quem era do outro campo viu e não comentou, porque a caixa de comentários punia quem discordava. Skinner (1957) afirma que o efeito de uma audiência negativa só se detecta em combinação com uma audiência positiva, como redução da força do comportamento apropriado a esta, e que, enquanto as audiências ficam separadas, o dilema não aparece; os fios de resposta mostram a consequência que esperava quem discordasse, mas não quantos deixaram de comentar por isso. A terceira é a filtragem: os 453 e 86 comentários que a API não devolveu, os filtros de comentários ofensivos e as pastas de mensagens ocultas do Direct podem ter retido justamente os ataques, e isso não foi conferido. A quarta é o conteúdo: imagens das câmeras e falas literais, no vídeo do 8 de janeiro, deixam menos a contestar do que uma série de preços, e o vídeo da gasolina, com dados que cada um confronta com a própria memória, teve nove vezes mais comentários do campo adversário. As quatro podem valer juntas, e o caso não as separa.

O caso dialoga com os estudos da seção 6.2 num ponto preciso. Nyhan et al. (2023) mostram que o conteúdo de fontes de mesma inclinação é a maior parte do que os usuários do Facebook veem, e que quem teve essa exposição reduzida engajou-se mais, por exposição, com o conteúdo congênial que ainda via. Aqui a exposição não veio da rede de quem segue quem, e mesmo assim a interação foi homofílica, no sentido de Cinelli et al. (2021): quem reagiu tinha, em esmagadora maioria, a inclinação do conteúdo. Que a homofilia das intera-

ções dependa menos da exposição do que das consequências que a audiência arranja é a hipótese do capítulo, e o caso é compatível com ela; a seleção pela entrega, a primeira rival acima, é igualmente compatível.

As cinco perguntas do Quadro 6.3 dizem até onde o caso vai. A correção existiu e é identificável: os comentários do campo adversário e as contestações sem campo corrigiam ou contestavam o vídeo, e a resposta em massa a eles é a consequência que a audiência arranhou para quem corrigia. A fala posterior não foi registrada: o caso é um retrato de uma noite, sem o antes e o depois de quem foi respondido, e não diz se essas pessoas voltaram a comentar. A audiência não é a mesma de outro momento, porque foi montada pela plataforma para estes vídeos. O reforço diferencial foi medido, nas curtidas e nas respostas. E a rival da aprovação sem esquiva continua de pé: as curtidas e as respostas bastam para descrever o que se viu, sem supor que alguém evitou a correção. Chamar o que aconteceu na caixa de comentários de esquiva de contracontrole, como o capítulo define o termo, exigiria a segunda pergunta, que o caso não responde.

<p>■ DADO</p>	<p>Dois vídeos de um perfil de 11.550 seguidores, publicados na campanha de 2026, somaram 2,02 milhões de visualizações; no mês, 99,7% das contas alcançadas pelo perfil não o seguiam, e o público alcançado era mais jovem e mais masculino que a base. Nos 2.952 comentários lidos, a declaração do campo adversário ao do vídeo foi de 1,5% (8 de janeiro) e 13,9% (gasolina). A adesão recebeu 91% das curtidas dadas a comentários no primeiro vídeo; no segundo, a contestação pelo campo de Lula recebeu 565 curtidas e a pelo campo de Bolsonaro, de mesmo tamanho, 49. Comentários do campo adversário receberam de 170 a 212 respostas por cem, e de 43 a 65 hostis; os do campo do vídeo, de 2 a 32 e de 0 a 2. Quatro comentários foram hostis ao autor. A API não devolveu 18% e 9% dos comentários contados.</p>
<p>■ INTERPRETAÇÃO</p>	<p>Os comentaristas funcionaram como uma subcomunidade verbal montada para cada vídeo, que reforçou a adesão e respondeu à discordância com consequências aversivas. A ausência de ataques ao autor é compatível com quatro leituras, a entrega seletiva, a audiência negativa, a filtragem pela plataforma e a forma do conteúdo, que o caso não separa.</p>

■ **HIPÓTESE** Se a resposta em massa à discordância funciona como punição, quem foi respondido comentará menos depois, e vídeos sobre temas em que cada lado confronta o dado com a própria experiência (preços) atrairão mais comentários do outro campo do que vídeos com imagens dos fatos. Testá-la pede o histórico de comentários das mesmas contas antes e depois, e a comparação com perfis de outros campos sob a mesma regra de codificação.

Para quem quiser repetir a análise num perfil, de qualquer campo, o Quadro 6.2 reúne os passos que o caso mostrou necessários.

QUADRO 6.2

Como medir a bolha de um perfil pelas consequências

1. Defina o período e uma linha de base: os posts habituais do perfil, com alcance e parcela de não seguidores, antes do episódio.
2. Leia, por post e para a conta, o alcance de seguidores e de não seguidores e o público por idade e gênero; anote a hora da leitura, porque os números mudam.
3. Colete todos os comentários e respostas pela API e registre quantos a plataforma conta e quantos devolve; antes de concluir que não houve ataques, confira no aplicativo os comentários ocultos, o filtro de palavras ocultas e as solicitações de mensagem ocultas.
4. Codifique posição, campo declarado e hostilidade com regras escritas (Quadro 6.1), e codifique à parte, com um segundo codificador, uma amostra sorteada, para medir a concordância.
5. Meça as consequências: curtidas por categoria de comentário, respostas e respostas hostis por campo de quem comentou, e os fios em que a discordância foi respondida.
6. Antes de chamar um repertório de esQUIVA, passe pelas cinco perguntas do Quadro 6.3; sem a fala posterior de quem foi respondido, o que se tem é descrição de consequências, e não de esQUIVA.

6.7 Análises semânticas em mídias sociais

Um dos livros do autor (Moreira et al., 2022a), que reúne também estudos com histórias em quadrinhos e com autoritarismo (Capítulo 4, seção 4.7), traz um estudo de análise semântica de postagens espontâneas, o Estudo 4. A pergunta era se, a partir de textos de rede social, seria possível descrever classes de equivalência ligadas ao movimento *Black Lives Matter* e, com elas, a visão social do movimento, medida pela influência semântica das palavras usadas. A coleta foi automática, por scripts em R que consultaram a API pública do Twitter durante 20 dias, de 22 de abril a 11 de maio de 2021, filtrada por hashtags do

movimento. Antes da análise, retiraram-se os nomes de usuário (de quem postou e de quem foi mencionado) e todas as palavras que não fossem nome próprio, adjetivo, substantivo ou verbo. As que restaram foram contadas e classificadas por classe gramatical e por influência semântica: positiva, negativa ou “não se aplica”; o capítulo não descreve o critério nem quem atribuiu a influência. Os resultados vêm em gráficos de frequência, sem números no texto: predominaram os substantivos, e os adjetivos foram poucos; a maior parte das palavras ficou em “não se aplica”, e, entre as positivas e as negativas, as negativas foram mais numerosas, inclusive entre os adjetivos; as palavras mais frequentes se relacionavam ao julgamento do caso George Floyd (“caso”, “júri”, “decisão”, “justiça”). Como a coleta ocorreu na semana do veredito desse julgamento, os autores concluem que não há como afirmar que a visão social do movimento é negativa, embora haja mais palavras de influência negativa que positiva, e sugerem repetir a coleta no momento da morte de Floyd. Registram dificuldades, como postagens incompletas e separação manual das palavras .

Os autores propõem, a partir daí, uma expansão conceitual: classes de equivalência “compartilhadas (ou comuns)”, encontradas em muitos indivíduos, que passam de fenômeno individual a fenômeno de alta frequência e se aproximam do conceito de representação social. Segundo os autores, coletar postagens espontâneas, sem que quem posta soubesse que seriam analisadas, exclui a variável de desejabilidade social, ao contrário dos questionários dos estudos 1 e 2 do mesmo livro.

Para este capítulo, a proposta importa por uma razão e exige uma ressalva. A razão é que, se a bolha é uma subcomunidade verbal, a hipótese natural é que as classes compartilhadas dentro dela sejam mais homogêneas que as da comunidade mais ampla, e o método do estudo, que mede a influência semântica das palavras num corpus de postagens espontâneas, é um modo de testá-la: bastaria separar o mesmo corpus por comunidade (as classes de equivalência estão no Capítulo 4, seção 4.2). A ressalva tem duas partes. O estudo não separou comunidades, porque os nomes de usuário foram retirados e a coleta foi por hashtag, e portanto nada diz sobre bolhas. E a espontaneidade da postagem não elimina a audiência: quem posta num feed fala a seguidores, e a audiência controla o que se posta, como no ciclo do Capítulo 9. O que o estudo exclui, a desejabilidade social diante de um pesquisador, é uma audiência, e não a ausência de audiência.

6.8 Interpretações rivais e limites

A leitura do capítulo, de que a bolha é uma subcomunidade verbal e de que parte do que nela se mantém pode ser esQUIVA de contracontrole, tem rivais. O quadro abaixo as reúne, uma a uma, e os limites do capítulo vêm depois dele.

Interpretações rivais para o mesmo conjunto de dados

Seleção, e não controle. A homofilia que Cinelli et al. (2021) medem pode refletir quem escolhe com quem falar, e não o que a comunidade faz com o que se diz: pessoas que já concordam se aproximam, e a bolha seria consequência da concordância, e não sua causa. Os estudos do capítulo não separam as duas direções.

Aprovação sem esQUIVA. Os repertórios da seção 6.4 podem ser mantidos só pelo reforço positivo da audiência, curtidas, repasses, pertencimento (Capítulo 9), sem que a correção seja aversiva ou que o repertório a evite. A aprovação basta para explicá-los. Distinguir as duas contingências exige mostrar que o repertório muda depois da correção, e nenhuma fonte do capítulo registra isso.

Desconfiança como discriminação, e não como esQUIVA. Quem desqualifica um corretor pode estar sob controle da história com aquela fonte. Freeze et al. (2021) mostraram, num experimento online, que avisos inválidos de desinformação reduzem a credibilidade de notícias autênticas e prejudicam a memória política. Participaram 434 pessoas de uma plataforma de crowdsourcing (MTurk) nos Estados Unidos, que viram um vídeo de discursos de deputados na C-SPAN, responderam a perguntas sem relação e leram um artigo noticioso fabricado sobre o vídeo, ora vago, ora com fatos corretos, ora com parte dos fatos alterados. Parte deles, sorteada, recebeu em seguida o aviso de que “alguma informação do artigo era imprecisa”, aviso inválido quando o artigo era exato; depois todos fizeram um teste de reconhecimento de 20 fatos do vídeo e avaliaram a credibilidade do vídeo e do artigo. O aviso fez o artigo parecer 0,59 ponto menos crível numa escala de 5, em todas as condições, inclusive quando era exato, e o efeito sobre a memória foi significativo em um de três testes. Se correções erram ou têm custos, desconfiar delas não é, por si, repertório de proteção, e a análise precisaria da taxa de acerto da fonte desqualificada, que o capítulo não tem.

Raciocínio motivado. Bail et al. (2018) propõem que a exposição a mensagens que conflitam com as atitudes provoca contra-argumentação e raciocínio motivado. É explicação dos autores, em termos cognitivos, e eles dizem que não identificaram o mecanismo. A leitura operante, de que o que o participante declara é controlado por audiência e história, é hipótese que o experimento não testa.

Migração por uso, e não por moderação. A Lupa lê a mudança de plataforma como migração para ambientes semiprivados e, ao mesmo tempo, avisa que seus resultados retratam o universo monitorado por ela (Agência Lupa, 2026). Evangelista e Bruno (2019) notam que o *zero-rating* favorece a popularidade do

WhatsApp. A presença maior do WhatsApp pode refletir o crescimento do uso da plataforma e do canal de recebimento da agência, e não uma fuga da moderação.

Amostras de um lado só. O material brasileiro cobre, em sua maior parte, grupos de um campo (Bonow Soares et al., 2021). O que se observa pode ser propriedade de qualquer grupo coeso e não deste, e o capítulo não tem como decidir.

Quatro limites valem para o capítulo inteiro. Primeiro, nenhuma fonte registra a correção e o que o falante fez depois dela, de modo que a esquiva de contracontrole continua hipótese em todos os casos. Segundo, exposição, interação e engajamento não são consequências arranjadas pela audiência, e os dados sobre câmaras de eco medem as primeiras. Terceiro, o material brasileiro é de uma plataforma principal, o WhatsApp, entre 2018 e 2022, de observação de poucos grupos, de grupos públicos monitorados e de um universo selecionado por uma agência; generalizar daí é extrapolação. Quarto, o termo “esquiva de contracontrole”, o agrupamento em quatro repertórios e as duas figuras são propostas aqui, e não de uma fonte.

Para quem quer usar o capítulo na análise de falas (Capítulos 3, 5 e 7), a consequência é prática. Afirmar que uma fala está protegida da correção exige o registro da correção e o da fala posterior, e uma checagem, por si, não traz esse segundo termo. O quadro abaixo reúne o que perguntar antes de chamar um repertório de esquiva.

QUADRO 6.3

Cinco perguntas antes de chamar um repertório de esquiva de contracontrole

1. Qual foi a *correção*: quem a fez, quando, com que rótulo, e ela chegou ao falante ou à comunidade?
2. A fala *posterior* mudou de forma, de fonte, de lugar ou de autoria em relação à anterior, e em quanto tempo?
3. A *audiência* é a mesma, com a mesma composição, antes e depois?
4. A audiência *reforçou* a mudança (reações, repasses, permanência) mais do que a forma anterior?
5. Que *rival* continua de pé: aprovação sem esquiva, discriminação da fonte, mudança de audiência ou de plataforma?

Síntese

A bolha, vista como subcomunidade verbal, é um conjunto de ouvintes com história de reforçamento comum. Os dados sobre câmaras de eco medem outra coisa: o que chega ao feed e com quem se interage. Em três estudos, a homofilia é grande em temas específicos, o feed inteiro é menos homogêneo do que o temor corrente supõe, e nem a redução da exposição congenial nem o aumento da exposição ao campo oposto deslocou as atitudes na direção esperada (seção 6.2). Nenhum deles mede quem reforça o que o falante diz nem o que da correção chega até ele, de modo que a pergunta da bolha como comunidade verbal continua aberta.

Guerin (1992) dá a condição sob a qual o dito contrafactual se mantém: um grupo pequeno e coeso, em que pertencer vale mais do que relatar com exatidão. Fora dela, a pressão de conformidade produz contracontrole (6.3). A proposta do capítulo é que há repertórios do falante e da comunidade que reduzem a chance ou a eficácia desse contracontrole, a esquiva de contracontrole, termo proposto aqui, de quatro tipos: forma, fonte, lugar e pessoas. Três enfraquecem condições que Delprato (2002) associa ao contracontrole eficaz; o quarto é acréscimo aqui proposto (6.4). O material brasileiro descreve cada um dos quatro, com limites de amostra, de plataforma e de campo político que a seção 6.5 enumera. O que ele não traz é a correção e o que o falante fez depois; por isso a esquiva continua hipótese, e a aprovação da audiência, sem esquiva, é a rival mais forte (6.7).

Os Capítulos 3 e 5 usam deste capítulo a ideia de que a forma da fala (atitude, autoclítico, mando disfarçado) pode ser mantida por reduzir a correção, e que isso se testa com falas datadas, antes e depois de checagens; o Capítulo 8, que o que a plataforma entrega ao feed decide quanto da fronteira da Figura 6.1 é permeável; o Capítulo 12, a pergunta de que intervenções agem sobre as contingências da comunidade, e não só sobre a exposição.

Referências

- Agência Lupa. (2026, 27 de julho). *Anatomia da desinformação eleitoral* [Relatório]. https://uploads.agencialupa.org/2026/07/RELATORIO-ANATOMIA-DA-DESINFORMACAO-ELEITORAL_LUPA-FINAL.pdf
- Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37), 9216–9221. <https://doi.org/10.1073/pnas.1804840115>

- Bonow Soares, F., Recuero, R., Volcan, T., Fagundes, G., & Sodré, G. (2021). Research note: Bolsonaro's firehose: How Covid-19 disinformation on WhatsApp was used to fight a government political crisis in Brazil. *Harvard Kennedy School Misinformation Review*, 2(1). <https://doi.org/10.37016/mr-2020-54>
- Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), Article e2023301118. <https://doi.org/10.1073/pnas.2023301118>
- Delprato, D. J. (2002). Countercontrol in behavior analysis. *The Behavior Analyst*, 25(2), 191–200. <https://doi.org/10.1007/bf03392057>
- Evangelista, R., & Bruno, F. (2019). WhatsApp and political instability in Brazil: Targeted messages and political radicalisation. *Internet Policy Review*, 8(4). <https://doi.org/10.14763/2019.4.1434>
- Freeze, M., Baumgartner, M., Bruno, P., Gunderson, J. R., Olin, J., Ross, M. Q., & Szafran, J. (2021). Fake claims of fake news: Political misinformation, warnings, and the tainted truth effect. *Political Behavior*, 43(4), 1433–1465. <https://doi.org/10.1007/s11109-020-09597-3>
- Guerin, B. (1992). Behavior analysis and the social construction of knowledge. *American Psychologist*, 47(11), 1423–1432. <https://doi.org/10.1037/0003-066x.47.11.1423>
- Guerin, B. (1994). Attitudes and beliefs as verbal behavior. *The Behavior Analyst*, 17(1), 155–163. <https://doi.org/10.1007/bf03392661>
- Moreira, M. B., Premuli, I. T., Rabelo, E. de P., Pessoa, M. B. F., & Medeiros, M. R. G. C. (2022a). *Atitudes: Histórias em quadrinhos, análises semânticas em mídias sociais e o paradigma de relações de equivalência de estímulos* (1ª ed.). Instituto Walden4. https://www.walden4.com.br/livros/pdfs/moreira_etal_2022_q0.pdf
- Nyhan, B., Settle, J., Thorson, E., Wojcieszak, M., Barberá, P., Chen, A. Y., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Guess, A. M., Kennedy, E., Kim, Y. M., Lazer, D., Malhotra, N., Moehler, D., . . . Tucker, J. A. (2023). Like-minded sources on Facebook are prevalent but not polarizing. *Nature*, 620(7972), 137–144. <https://doi.org/10.1038/s41586-023-06297-w>
- Recuero, R., Tavares, C., Volcan, T., Pozzebon, M., & Dutra, M. (2025). Desinformação narrada pela checagem: estudo das eleições de 2018 e 2022. *E-Compós*, 28, 1–26. <https://doi.org/10.30962/ecomps.3039>
- Ricard, J., & Medeiros, J. (2020). Using misinformation as a political weapon: COVID-19 and Bolsonaro in Brazil. *Harvard Kennedy School Misinformation Review*, 1(2). <https://doi.org/10.37016/mr-2020-013>
- Skinner, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts.



LAB WALDEN4

Ref faça os experimentos clássicos no navegador

Estudos clássicos da psicologia reconstruídos para rodar no computador ou no celular. Você participa, e os seus dados aparecem ao lado dos resultados originais.

8 experimentos para fazer

Acesse, de graça e sem cadastro

walden4.com.br/experimentos



7

O voto como comportamento verbal

Um candidato pode falar com os eleitores até a véspera da eleição, mas é proibido de entrar na cabine de votação com eles. A observação é de Lloyd (1994), que a faz para ilustrar outro problema, o de que o agente de mudança pode ter acesso apenas ao comportamento verbal e nunca ao comportamento que lhe interessa. Serve de abertura aqui porque nomeia a assimetria que organiza o capítulo: o que o eleitor diz tem audiência, e o voto, por ser secreto (Visser, 1996), não a tem no momento em que ocorre.

Os Capítulos 3 e 5 analisaram o que o político diz. Este capítulo volta-se para quem o ouve e vota. A pergunta é por que votar pode ser tratado como comportamento verbal, isto é, como comportamento mantido por consequências que outras pessoas mediam, e o que isso muda na leitura da relação entre o que o eleitor diz e o que ele faz na urna. Os instrumentos são os do Capítulo 3, o operante verbal e a audiência, e os do Capítulo 4, a atitude como comportamento verbal avaliativo e a transferência de função entre estímulos de uma mesma classe. Falta um, que entra na seção 7.4, no ponto em que a análise precisa dele: a correspondência entre dizer e fazer.

Um aviso sobre as fontes. Nenhum dos estudos usados aqui mede o voto de eleitores brasileiros. Visser (1996) analisa uma pesquisa eleitoral holandesa; Bond et al. (2012), uma eleição legislativa dos Estados Unidos; os estudos de atitude e comportamento são, em sua maior parte, de estudantes estado-unidenses, entre as décadas de 1930 e 1970. Os estudos brasileiros disponíveis, os do Capítulo 4, tratam de atitudes diante de presidentes, e não de votos. O capítulo é, portanto, uma proposta de análise, e a extensão dos achados ao Brasil é hipótese, não relato do que se sabe do voto brasileiro.

7.1 O voto pela definição de comportamento verbal

Para Skinner (1957), comportamento verbal é o comportamento reforçado pela mediação de outras pessoas (o ouvinte), já o conjunto dos comportamentos de falante e ouvinte é o “episódio verbal total”, em que nada “emerge” na unidade social além do comportamento combinado de dois ou mais indivíduos. A definição não especifica forma, modo nem meio: qualquer movimento capaz de afetar outro organismo pode ser verbal. Apontar é verbal, porque só é eficaz quando altera o comportamento de alguém.

A aplicação ao voto é feita aqui, e convém dizer onde ela é frágil. Marcar um voto é um movimento que tem efeito sobre um meio, a cédula ou a tela, e por meio dele sobre o resultado da eleição, sobre quem governa e sobre o que a comunidade do eleitor faz em seguida. Pelo critério de Skinner, é verbal na medida em que a consequência que o mantém é mediada por pessoas. A fragilidade está em que o ouvinte desse comportamento não é uma pessoa presente que reforça o ato, como é o ouvinte de um pedido; é um sistema, e o reforço do ato, no sentido em que o definimos para o comportamento verbal, não ocorre no momento em que ele é emitido. O capítulo não resolve a classificação por decreto. Adota uma afirmação mais modesta: o voto é comportamento cuja manutenção é social, e por isso cabe a ele a análise de relações de controle mediadas, porque o que o mantém é o que a comunidade do eleitor faz com ele e com o que o eleitor diz dele.

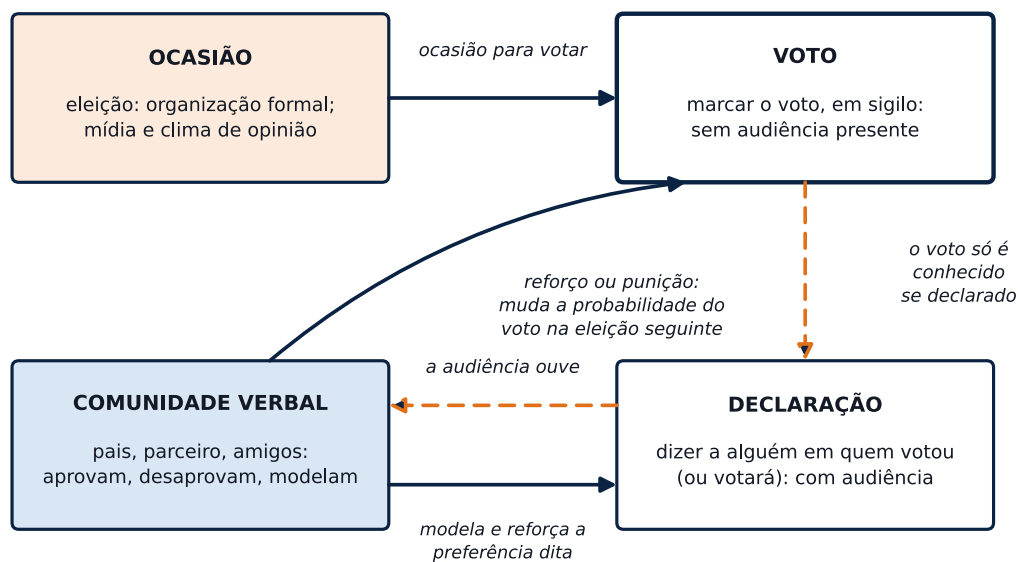
Visser (1996), cujo artigo serve de base à seção, chega à mesma afirmação modesta por outro caminho. O autor afirma que o voto foi raramente estudado pela perspectiva analítico-comportamental, em razão da forte influência mentalista na pesquisa eleitoral. Parte de que votar num partido é uma resposta operante, emitida sob condições de estímulo especificadas: a organização formal da eleição é a ocasião; as consequências são aplicadas pela comunidade verbal do eleitor (família, amigos, colegas, vizinhos); e fontes distais, como o clima da opinião pública, que a mídia torna relevante para cada eleitor, também especificam probabilidades de reforço ou de desaprovação.

Para testar o modelo, Visser o simplifica. Deixa de lado os estímulos discriminativos da situação eleitoral, supondo que basta o eleitor saber quando e onde ocorre a eleição e quais partidos concorrem, e fica com dois agentes de consequências: um ambiente social próximo (como família), que modelam o comportamento e reforçam a verbalização de preferência igual à deles, e, na vida adulta, o pares (comunidade verbal mais ampliada: amigos, colegas de traba-

lho, parceiros), que reforça o voto semelhante ao seu e desaprova o divergente. A Figura 7.1 reúne o modelo e acrescenta uma caixa que ele não traz, a declaração.

FIGURA 7.1

O voto como episódio verbal mediado: ocasião, voto sem audiência, declaração com audiência e comunidade verbal



Contínua: relação do modelo de Visser (1996). Tracejada: leitura aqui proposta (a declaração não está no modelo).

Nota. Esquema original do autor, a partir de Visser (1996) e de Skinner (1957). Não representa dados. Linha contínua: relação do modelo de Visser; tracejada: leitura aqui proposta (a declaração não consta do modelo).

A caixa da declaração é o acréscimo que torna a análise possível, e Visser não o faz. No modelo, as consequências sociais incidem sobre o voto. Mas o voto é secreto, e o que a comunidade do eleitor pode reforçar ou punir é o que ela vê e ouve: o que o eleitor diz sobre seu voto, antes ou depois, e o que se pode inferir do que ele diz de outros assuntos. Entre o ato e a consequência social há, na maior parte dos casos, um enunciado. É por esse enunciado que o voto entra no episódio verbal, e é nele que o resto do capítulo trabalha.

7.2 Mando, tato ou intraverbal? Funções possíveis do voto

Como no Capítulo 3, o que define o operante é o que o controla, e não a forma. Skinner (1957) afirma que o mando é um tipo de operante definido por suas variáveis de controle, e não uma unidade formal: nenhuma resposta pode ser

dita mando só pela forma, e identificar qualquer operante verbal exige conhecer as variáveis de que a resposta é função. O mesmo vale para o tato e para o intraverbal. Um mesmo voto, no mesmo candidato, pode estar sob controles diferentes em eleitores diferentes, e no mesmo eleitor em eleições diferentes.

A primeira candidata é a função de tato. O tato é o operante evocado por um objeto, evento ou propriedade, porque na presença deles uma resposta dessa forma é caracteristicamente reforçada na comunidade verbal (Skinner, 1957). Um voto sob essa função seria controlado pelo que o candidato ou o partido fez, disse ou representa para o eleitor, e variaria com o evento: se o que o controla mudasse, o voto mudaria. É a leitura mais direta do voto. Seu ponto fraco é que ela exige contato com o evento. Guerin (1994) observa que muitas atitudes versam sobre eventos que a pessoa nunca experienciou, como a energia nuclear, e por isso seriam intraverbais: o que se sabe vem de jornais e de outras pessoas. É razoável supor que, para muitos eleitores, candidatos a cargos nacionais sejam desse tipo, mas não há dado sobre isso entre as fontes lidas.

A segunda é a função de intraverbal, a resposta sem correspondência ponto a ponto com o estímulo verbal que a evoca, de que Skinner (1957, p. 71–72) diz que a maior parte dos “fatos” da história é feita. Um voto sob essa função seria controlado por palavras: a sigla, o slogan, o que a família diz, o que o grupo recomenda. O modelo de Visser é, nesse ponto, quase uma descrição dessa história: a audiência reforça a verbalização da preferência semelhante à dela, de modo que, quando a eleição chega, o eleitor tem uma preferência verbalizada que o ato apenas continua. Nesse caso, o voto muda quando mudam as palavras que o evocam e as pessoas que as dizem, e não quando muda o evento.

A terceira é a função de mando, ou melhor, de algo que se aproxima dele. O mando é a resposta reforçada por uma consequência característica e, por isso, sob controle de privação ou de estimulação aversiva (Skinner, 1957). Votar para obter ou para evitar algo de quem governa é a leitura mais natural, e a mais frágil: o ouvinte não está presente, não foi treinado a reforçar o eleitor, e o reforçador, se vem, vem à comunidade inteira e com grande atraso. A função de mando cabe melhor na declaração do que no voto. Guerin (1994) mostra o mecanismo para enunciados de atitude: a chefe que diz não gostar de quem toma café no trabalho não faz um tato inocente de atitude privada, e funciona como se tivesse mandado “Não tome café enquanto trabalha, ou então!”, tenha ou não essa intenção. Quem diz “eu voto em fulano” diante de alguém que ainda não decidiu pode estar fazendo isso, e a regra do Capítulo 5 vale aqui: a forma não decide.

Há uma quarta possibilidade, que não é um operante verbal a mais mas muda o que se pode esperar do voto: a regra. Street (1994), citando um texto de Craig, de 1980, dá o exemplo do eleitor que diz ser democrata e vota num republicano para presidente. A pesquisa tradicional toma isso por inconsistência de atitude; para Craig, pode ser apenas artefato da incapacidade do pesquisador de identificar aspectos salientes do estímulo, dada a história da pessoa, porque o eleitor pode ter uma regra de nunca votar em presidente do partido que controla o Congresso. A “inconsistência” é, então, medida da incapacidade do observador de fazer sondas verbais eficazes, dada a história do falante. O voto que decorre de uma regra é comportamento governado por regra, nos termos da seção 3.7, e pode não ser nem tato, nem intraverbal, nem mando.

O que se pode dizer, no estado das fontes, é pouco. As quatro possibilidades são candidatas e nenhuma foi testada para o voto de um eleitor brasileiro. O único modelo operante de voto entre as fontes é o de Visser, que trata o voto como resposta mantida por consequências sociais e deixa de lado os estímulos discriminativos da eleição. O que a análise ganha com a lista é uma pergunta precisa para cada caso: o que mudaria o voto deste eleitor, e quem o saberia.

7.3 O voto secreto como retirada da audiência

A audiência é, para Skinner (1957), o estímulo discriminativo em cuja presença o comportamento verbal é caracteristicamente reforçado. As audiências diferem pelo que fazem com o comportamento verbal: a que apenas não reforça se distingue da que pune, e uma audiência em cuja presença o comportamento verbal é punido é uma audiência negativa. O confidente é o exemplo da audiência não punitiva, a pessoa simpática a quem se pode falar com menos medo de punição do que a ouvintes ao acaso, e, diante dela, o falante pode emitir comportamento que outras audiências punem. O Capítulo 10 (seção 10.1) trata dessas audiências por inteiro; aqui bastam duas ideias: a audiência é parte da ocasião do comportamento verbal, e o que uma audiência pune pode ser emitido diante de outra.

O sigilo da cabine é, nesses termos, a retirada institucional de qualquer audiência identificada no momento do ato. O eleitor marca o voto sem que ninguém que o aprove ou o reprove esteja presente, e sem que o ato, no momento, produza consequência social. A hipótese do capítulo, que nenhuma fonte verificada testa, é que isso desloca o controle: a audiência deixa de ser ocasião para o ato, e o que sobra é o que a história do eleitor deixou, inclusive regras e preferências verbalizadas, aprendidas diante dessas audiências. O que a reti-

rada não elimina é o que o eleitor disse antes e o que dirá depois. A declaração de voto, ao contrário do voto, é emitida diante de uma audiência, e é sob o controle dela que se espera encontrá-la.

Há uma consequência para quem pesquisa, que Visser (1996) anota com franqueza. Como o voto é secreto, o pesquisador não o pode inspecionar e recorre a relatos verbais do próprio eleitor sobre seu comportamento de votar e sobre os fatores que o reforçam. Esses relatos podem estar sob controle de contingências diferentes das do voto que supostamente refletem, em particular as da situação de entrevista em que os dados de pesquisa costumam ser colhidos. Visser em seguida supõe que, numa entrevista de pesquisa que procura ser discreta, o relato reflete com precisão o comportamento de votar e os fatores que o controlam. É uma suposição de trabalho, declarada como tal, e o resto do artigo depende dela.

Guerin (1994) fornece o contraponto. Escalas e perguntas de pesquisa dependem de quem pergunta e da cultura verbal em que a pesquisa é feita, porque atitudes dependem sempre de consequências sociais; e cita, nesse ponto, um estudo de 1986 sobre a presença de outros e a declaração do voto nas eleições americanas. Esse estudo não foi lido; só se conhece a referência, e não se afirma o que ele achou. Para o capítulo, basta o que Guerin extrai: o relato de voto, como outro relato, depende de quem pergunta.

■ DADO	Como o voto é secreto, o pesquisador não o inspeciona e usa relatos verbais, que podem estar sob controle de contingências diferentes das do voto, em particular as da entrevista (Visser, 1996). Escalas e perguntas de pesquisa dependem de quem pergunta e da cultura verbal em que a pesquisa é feita (Guerin, 1994).
■ INTERPRETAÇÃO	O que a pesquisa eleitoral observa é a declaração do eleitor sob a audiência da entrevista, e não o voto. A relação entre os dois é o objeto de estudo, e não pode ser pressuposta.
■ HIPÓTESE	Que a declaração pública de voto esteja mais sob controle da audiência do que o voto está é hipótese razoável e não testada: não há, entre as fontes verificadas, estudo que compare, para os mesmos eleitores, o que declaram diante de audiências diferentes com o voto validado.

7.4 Declarar e votar: atitudes verbais e atos

A pergunta pela relação entre o que a pessoa diz e o que ela faz é mais velha que a pesquisa eleitoral, e a análise do comportamento não é a única a tê-la feito. Lloyd (1994) parte do reconhecimento de senso comum de que há correlação baixa, ou nula, entre o que a pessoa diz que faria e o que de fato faz depois, e nota que o psicólogo, que se apoia com frequência no relato verbal em vez de observar o comportamento critério, é menos cético que o leigo. A seção reúne o que as fontes verificadas mostram, e começa pelo conceito que a organiza.

CONCEITO EM USO

Correspondência dizer-fazer

Correspondência dizer-fazer é a relação entre o que a pessoa diz que fez, ou que fará, e o que de fato faz. Em Risley e Hart (1968), doze crianças de 4 a 5 anos, de famílias de baixa renda, numa pré-escola experimental de meio período, formavam dois grupos de seis. As professoras registravam o uso dos materiais na hora livre da manhã (o fazer) e, cerca de uma hora e meia depois, no último lanche antes de as crianças irem para casa, perguntavam a cada uma o que tinha feito de bom naquele dia (o dizer). Primeiro, o lanche e a aprovação vinham de relatar o uso de um material específico: blocos no grupo A, tinta no grupo B. Com o lanche dependendo só do relato, a proporção de crianças que relatavam subiu a uma mediana de 100% nos dois grupos, sem efeito confiável sobre o uso. Depois, o lanche passou a depender de que o relato correspondesse ao uso observado, e a correspondência subiu, por caminhos diferentes: no grupo A, o uso aumentou até o nível do relato (mediana de 100% de dizer e de fazer); no grupo B, o relato caiu (a uma mediana de 40%) até o nível de quem de fato pintou (25%).

Duas consequências importam aqui. A primeira é que dizer e fazer são, nas palavras de Lloyd (1994), classes de respostas separadas, cujo desenvolvimento pode ou não ser independente, e não há implicação de que uma cause ou influencie a outra. A segunda é que a correspondência não é propriedade da pessoa: é efeito de uma contingência que a reforça. Reforçar o dizer sozinho não produz o fazer.

Risley e Hart partiram de estudos anteriores que reforçaram só o comportamento verbal e encontraram pouco aumento do comportamento não verbal relacionado. O mesmo estudo abre, em Lloyd (1994), a seção sobre fazer-dizer e dizer-fazer, e é por ele que o conceito entra na discussão de atitudes. A Tabela 7.1 reúne cinco fontes verificadas: duas, que o texto conhece pelo relato de

Lloyd, mostram a distância num estudo de campo e numa revisão de 31 estudos; duas a medem num pedido concreto feito ao participante; e uma mostra uma condição que a reduz.

TABELA 7.1

Cinco fontes sobre a distância entre o que se diz e o que se faz, e uma condição que a reduz

Fonte	“Dizer”	“Fazer”	Resultado	O que não permite
LaPiere, em 1934, pelo relato de Lloyd (1994)	Resposta escrita de proprietários sobre servir clientes chineses	Atendimento de um casal chinês, em visita a cerca de 250 estabelecimentos	Cerca de 90% dos 128 que responderam disseram que não serviriam	Conferir o estudo original, que não foi lido
Wicker, em 1969, revisão relatada por Lloyd (1994)	Escalas de atitude	Medidas de comportamento aberto, em 31 estudos	Correlações raramente acima de +0,30, muitas perto de zero, negativas em voto quatro	Conferir os 31 estudos; falar de
DeFleur e Westie (1958)	Escala de diferenças somadas; 23 estudantes mais e 23 menos preconceituosos, pareados	Assinar, num formulário de usos graduados, a permissão de ser fotografado com pessoa negra do sexo oposto	Preconceituosos evitaram mais, mas 14 de 46 fizeram o oposto do previsto; 71,8% invocaram grupo de referência	Generalizar: 46 estudantes; os autores não testaram o mecanismo
Linn (1965)	Atitude verbal (preconceito)	Posar para fotografia com pessoa negra do sexo oposto	Discrepância em 59% dos casos; relação função do envolvimento social e da experiência prévia	Estimar a taxa em outras populações
Regan e Fazio (1977)	Atitude, formada por experiência direta ou por informação	Comportamento relacionado, em laboratório e em campo (crise de moradia)	Mais consistência com experiência direta	Falar de declarações públicas; de voto

Nota. Números dos artigos e, nas duas primeiras linhas, do relato de Lloyd (1994). Os estudos de LaPiere e de Wicker são conhecidos aqui só por esse relato e, no caso de LaPiere, por menção em DeFleur e Westie (1958); as páginas citadas são as de Lloyd. “Medidas de comportamento” em Wicker são as que os 31 estudos usaram, e não o voto.

As duas primeiras linhas dizem o que a pergunta tem de velha. LaPiere viajou pelos Estados Unidos com um casal chinês, que foi atendido em mais de 250 restaurantes, hotéis e estabelecimentos semelhantes, com uma única recusa (DeFleur e Westie, 1958); quando perguntou depois, por escrito, se aceitariam clientes chineses, cerca de 90% dos 128 proprietários que responderam disseram que não (Lloyd, 1994). Wicker revisou 31 estudos que mediram atitude por escalas e comportamento aberto: as correlações raramente passaram de +0,30, muitas ficaram perto de zero e em quatro estudos foram negativas

(Lloyd, 1994). Uma correlação de +0,30 é fraca: elevada ao quadrado, indica que menos de um décimo da variação de uma medida acompanha a da outra. O número que importa para o capítulo é o de Wicker: a distância entre o dizer e o fazer não é exceção. A seriedade do número, escreve Lloyd, é maior quando se nota a aceitação difundida, entre leigos e psicólogos, da ideia de que, para mudar o comportamento, deve-se antes mudar atitudes, ideia que exige uma correlação alta e positiva.

As duas linhas seguintes são estudos em que o fazer foi uma decisão diante de um pedido concreto, e o que interessa é o que os autores dizem da origem da discrepância. DeFleur e Westie (1958) aplicaram escalas de preconceito a 250 estudantes e escolheram 23 do quartil mais preconceituoso e 23 do menos preconceituoso, pareados em oito características. Depois de uma sessão de laboratório e de uma entrevista, pediram que cada um assinasse, num formulário de autorização com uma série graduada de usos, a permissão de ser fotografado com uma pessoa negra do sexo oposto. Os preconceituosos evitaram mais, mas 14 dos 46 mostraram comportamento oposto à atitude verbal, e a relação, embora significativa, não foi de um para um. Em 71,8% dos casos os sujeitos invocaram algum grupo de referência ao decidir assinar ou recusar, e todos os que o fizeram mencionaram algum grupo de pares; os autores atribuíram a falta de relação linear a restrições sociais e sugeriram que as escalas de atitude talvez precisem de uma categorização sistemática delas.

Linn (1965) refez o desenho com a medida de atitude idêntica ao ato. Trinta e quatro universitárias de Wisconsin, de 18 e 19 anos, tinham respondido a um questionário com sete itens do tipo “eu posaria para uma fotografia com uma pessoa negra do sexo oposto se a imagem fosse usada em tal contexto”, de uso em laboratório a folheto de campanha nacional pela integração racial. Quatro semanas depois, sem saber que participavam de um estudo, foram convidadas a ajudar a construir um teste de personalidade e assinaram ou não as mesmas sete autorizações de uso. Em 59% dos casos, a atitude verbal e o número de autorizações assinadas diferiram em dois ou mais pontos de sete. Linn concluiu que a relação entre preconceito e discriminação é função do nível de envolvimento social com o objeto e da experiência prévia com ele. Em linguagem operante, restrições sociais e grupos de referência são audiência (seções 7.1 a 7.3).

A condição que reduz a distância, por sua vez, é a experiência direta. Regan e Fazio (1977) fizeram um experimento de laboratório e um estudo de campo. No laboratório, 28 calouros voluntários da Cornell foram sorteados para duas condições: na de experiência direta, resolveram os exemplos de cinco tipos de problema; na de experiência indireta, viram os mesmos exemplos já resolvidos e explicados pelo experimentador. Todos avaliaram o interesse de cada

tipo, de “extremamente chato” a “extremamente interessante”, o que mediu a atitude, e depois tiveram 15 minutos livres para trabalhar os problemas que quisessem; a ordem e a proporção tentada de cada tipo mediram o comportamento. A consistência entre atitude e comportamento foi maior nos sujeitos cuja atitude se formou por experiência comportamental direta do que nos de atitude formada por informação do experimentador. No estudo de campo, questionários foram enviados a calouros que a universidade alojara em moradia temporária por falta de vagas (experiência direta da crise de moradia) e a calouros alojados desde o início em moradia permanente (experiência indireta); mediram a atitude sobre a crise em sete itens e ofereceram seis ações possíveis, como assinar uma petição ou escrever uma carta, e a consistência entre atitude e ação foi maior nos estudantes com experiência direta. Para o voto, a pergunta é qual seria a experiência direta com o objeto: o candidato chega, para muitos eleitores, por palavras (seção 7.2), e a experiência direta que o eleitor tem é a de votar e a de declarar o voto.

Para os analistas do comportamento, o problema não está só na medida. Street (1994) escreve que, na visão analítico-comportamental, atitudes são os comportamentos que se veem quando se pede a alguém que preencha uma pesquisa de atitude, sem estado mental duradouro sendo captado, e que responder a uma escala é comportamento emitido sob condições sociais incomuns. O cognitivista supõe que autorrelato e comportamento diante do objeto são efeitos da mesma causa, a atitude; o analista, que cada um é produzido por seus próprios eventos causais e que, se são diferentes, os comportamentos provavelmente serão diferentes. O critério de consistência é, ele próprio, verbal, e os autorrelatos são reforçados socialmente por plausibilidade, e não por exatidão objetiva. Guerin (1994) chega ao mesmo lugar: dadas as múltiplas fontes de controle, os repertórios amplos e as várias comunidades verbais, devem-se esperar inconsistências entre atitude e comportamento, sobretudo quando o comportamento é verbal.

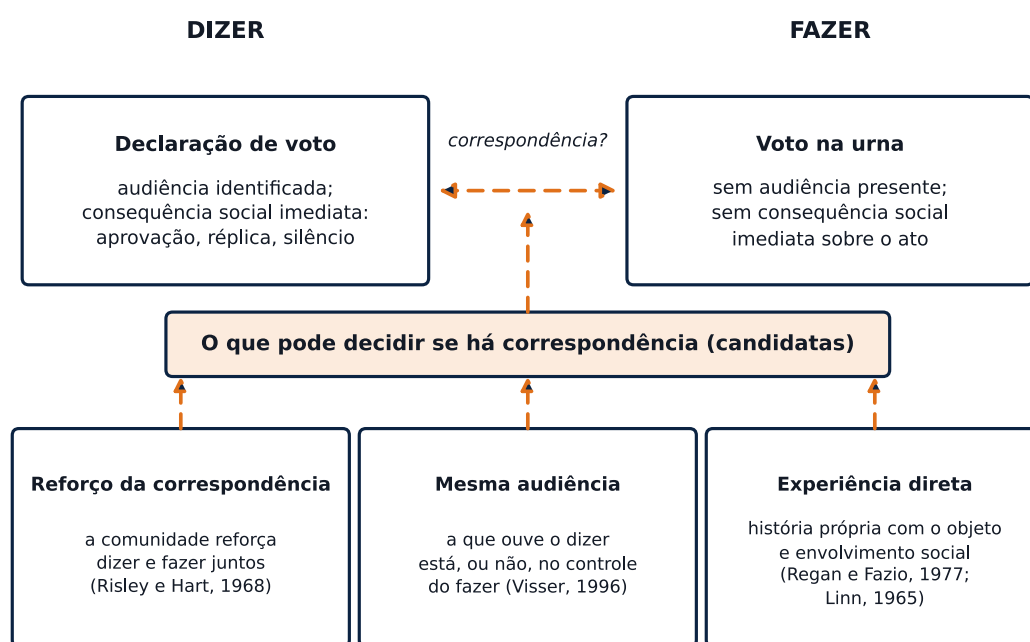
A Figura 7.2 reduz a seção a uma pergunta: o que decide se o que o eleitor diz e o que ele faz correspondem? As três candidatas saem das fontes: a história de reforço da correspondência (Risley e Hart, 1968); a identidade, ou não, entre a audiência do dizer e a que controla o fazer, leitura que decorre de Visser (1996) e de DeFleur e Westie (1958); e a experiência direta com o objeto (Regan e Fazio, 1977; Linn, 1965). Nenhuma foi testada para declaração de voto e voto no mesmo eleitor.

Lloyd (1994) acrescenta duas observações sobre a ordem, ambas de estudos que não foram lidos. Num estudo de Israel e O’Leary, de 1973, crianças que brincavam por 20 minutos e, 40 minutos depois, diziam o que haviam brincado (fazer-dizer) mostraram maior correspondência que as que diziam o que

brincariam e, 40 minutos depois, brincavam (dizer-fazer). E, num estudo de Sowers e colaboradores, de 1977, o relato de universitários sobre o estudo feito 14 horas antes correspondeu mais ao estudo do que o planejamento deles correspondeu ao estudo feito 6 horas depois. A leitura do capítulo é que declarar depois do voto, como numa declaração colhida na saída da urna, é uma resposta do tipo fazer-dizer, e declarar antes é do tipo dizer-fazer.

FIGURA 7.2

Dizer e fazer: o que pode decidir se a declaração de voto e o voto correspondem



Esquema original, sem dados. Tracejada: relação conceitual ou hipótese.

Nota. Esquema original do autor, a partir de Risley e Hart (1968), Visser (1996), Regan e Fazio (1977) e Linn (1965). Não representa dados. Linha tracejada: relação conceitual ou hipótese.

■ **DADO**

Em 31 estudos revistos por Wicker em 1969, as correlações entre escalas de atitude e medidas de comportamento raramente passaram de +0,30 (Lloyd, 1994). Em DeFleur e Westie (1958), 14 de 46 estudantes agiram em oposição à atitude verbal; em Linn (1965), a discrepância foi de 59%. Em Risley e Hart (1968), reforçar o relato elevou o relato sem alterar o uso, e reforçar a correspondência a produziu nos dois grupos, por caminhos diferentes.

■ INTERPRETAÇÃO

A distância entre o dizer e o fazer não é anomalia a explicar por erro de medida: é o que se espera de duas classes de respostas com contingências próprias (Street, 1994; Guerin, 1994). Quando há correspondência, parece depender de uma história de reforço da própria correspondência (Risley e Hart, 1968) e de experiência direta com o objeto (Regan e Fazio, 1977). Nada nas fontes é sobre voto.

■ HIPÓTESE

Que a correspondência entre declaração de voto e voto seja maior quando a mesma audiência participa dos dois momentos, e menor quando a audiência da declaração difere da que controla o voto, é hipótese aqui proposta, e nenhuma fonte a testa. Que declarar depois de votar corresponda mais do que declarar antes, como Lloyd relata para crianças e universitários, é testável com votos validados e declarações colhidas em momentos diferentes, e não há, entre as fontes lidas, estudo que o faça.

7.5 Consequências sociais do votar: os experimentos de campo

Se a análise do voto como comportamento verbal está certa, o que a comunidade do eleitor faz com o voto, ou com o que se diz dele, muda a probabilidade de votar e de votar do mesmo modo. Duas fontes tratam disso, de desenhos opostos: um levantamento, que só pode associar, e um experimento de campo, que manipulou.

O levantamento é o de Visser (1996), com dados do estudo eleitoral nacional holandês de 1986, uma pesquisa em duas ondas, antes e logo depois da eleição, e a única da série que mediu o voto das pessoas próximas do eleitor. Visser classifica cada eleitor segundo a história de reforço que o modelo supõe, isto é, segundo se os pais votam no mesmo partido e se há um parceiro cujo partido se conhece, e verifica se o voto do eleitor coincide com o do parceiro ou, sem parceiro identificável, com o dos pais; nesse caso, a direção do voto é considerada explicada pelo modelo. Entre 1.255 eleitores, o modelo explica a direção de 812 votos (65%), contra 41% esperados caso os eleitores fossem distribuídos ao acaso entre as categorias. Quanto à consistência, que Visser define como o número de vezes em que o eleitor escolhe o mesmo partido em eleições consecutivas, os votos modelados ou reforçados por pais ou parceiro foram mais consistentes que os desaprovados por pessoas significativas, em geral com diferença grande e significativa: 68% (pais e parceiro com o mesmo voto) e 67% (só pais) dos eleitores disseram ter votado sempre no mesmo partido, contra 29% a 36% nas células em que o eleitor divergia de pais ou parceiro. A exceção que o autor destaca é a célula em que pais e parceiro votavam di-

ferente e o eleitor votava como o parceiro: 32%, o que, para ele, sublinha a importância da história de reforço. Visser ressalva que a pesquisa por levantamento carece do caráter controlado das experimentais, de modo que afirmações causais exigem cautela, e que a análise ao nível do eleitor individual pode deixar de ver o impacto de desenvolvimentos macrossociais; e conclui que, do ponto de vista analítico-comportamental, não há razões válidas para conceder ao voto posição privilegiada entre as ações humanas.

O experimento é o de Bond et al. (2012), com 61 milhões de usuários do Facebook nas eleições legislativas americanas de 2010. Os usuários foram designados ao acaso a uma mensagem social (60.055.176 usuários: incentivo a votar, link para o local de votação, botão “I Voted”, contador e até seis fotos de amigos que haviam clicado no botão), a uma mensagem informativa (611.044: o mesmo, sem as fotos) ou a nenhuma mensagem (613.096). Mediram três coisas. Clicar em “I Voted” é medida de autoexpressão política, que os autores dizem ser provavelmente afetada pelo quanto o usuário deseja ser visto como eleitor pelos outros; clicar no link do local de votação mede busca de informação, e não é relatado à comunidade social; e o voto foi validado em registros públicos para 6,3 milhões de usuários.

Os resultados são três. Os que receberam a mensagem social clicaram mais em “I Voted” que os da informativa (20,04% contra 17,96%), e votaram 0,39% mais que os que não receberam mensagem e 0,39% mais que os da informativa; o comparecimento dos que receberam a informativa foi idêntico ao dos controles. Para cada amigo próximo que recebeu a mensagem social, o usuário ficou 0,224% mais propenso a votar (voto validado) e 0,099% a expressar voto; não houve evidência de efeito de amigos comuns sobre o voto validado nem sobre a busca de local. E os autores estimam que a mensagem social aumentou o comparecimento diretamente em cerca de 60 mil eleitores e, por contágio social, em outros 280 mil, num total de 340 mil votos, cerca de 0,14% da população em idade de votar; todos os efeitos são de intenção de tratar, isto é, calculados sobre todos os designados à mensagem, inclusive quem já havia votado antecipadamente ou entrou tarde demais no Facebook para votar; por isso subestimam o efeito sobre quem de fato pôde ser tratado. Quase toda a transmissão ocorreu entre amigos próximos, que provavelmente têm relação face a face.

Em termos operantes, a mensagem social difere da informativa por uma única coisa: as fotos de amigos que declararam ter votado. O experimento isola, portanto, que o dado social, e não a informação sobre onde e quando votar, acrescentou comparecimento. O que, no dado social, atuou, o desenho não separa. Três leituras são compatíveis: a de modelo (ver amigos votando); a de audiência (a antecipação de que os amigos verão o próprio “I Voted”, a autoexpressão

que os autores medem); e a de regra (ver que pessoas do grupo votam torna aplicável uma regra que o eleitor já tinha). O dizer e o fazer foram medidos sob a mesma manipulação e variaram juntos, no sentido de que a mensagem social aumentou os cliques e o voto validado. O que o artigo não mostra é a correspondência por usuário: o voto validado foi estudado por um processo em nível de grupo, descrito no material suplementar, e não se pode ver, para cada pessoa, se o clique correspondeu ao voto.

Há outros experimentos de campo sobre pressão social e comparecimento, e eles não são citados porque o texto integral não pôde ser lido por via legal. A lacuna é do acervo de fontes lidas, e não da literatura. Também por isso o capítulo não afirma que o efeito de Bond et al. se estenda a eleições de outros países, a outras plataformas ou a outros anos.

<p>■ DADO</p>	<p>Em 1.255 eleitores holandeses, os votos modelados ou reforçados por pais ou parceiro foram mais consistentes ao longo das eleições que os desaprovados, e a direção do voto foi explicada em 65% dos casos, contra 41% esperados por acaso (Visser, 1996). Em 61 milhões de usuários americanos, a mensagem com fotos de amigos aumentou o clique em “I Voted” e o voto validado (0,39% a mais que sem mensagem); a mensagem sem fotos não alterou o comparecimento (Bond et al., 2012).</p>
<p>■ INTERPRETAÇÃO</p>	<p>Os dois são compatíveis com a leitura de que o que a comunidade do eleitor faz com o voto, ou com o que se diz dele, muda a probabilidade de votar e de votar do mesmo modo. O de Bond et al. é de manipulação e de uma plataforma; o de Visser é de associação, e o próprio autor adverte sobre a causalidade.</p>
<p>■ HIPÓTESE</p>	<p>Que o mecanismo seja a aprovação ou a desaprovação antecipada pelos amigos (reforço social) é hipótese, e o desenho não a separa de modelo e de regra. Que o mesmo valha para eleições brasileiras é extensão, e não dado.</p>

7.6 Nomes e números de candidatos em classes de equivalência

Nomes, siglas e números de candidatos são os estímulos com que o eleitor chega à urna, e o Capítulo 4 (seção 4.5) deixou uma hipótese sobre eles: que o rótulo direita-esquerda faça parte da classe com que o nome do político chega ao eleitor. O que o Capítulo 4 tinha para sustentá-la é pouco. No estudo de Pereira e Moreira (2021b), cinco estudantes de Psicologia aprenderam, por tentativa e erro, quais figuras abstratas combinavam com cada uma de três fotografias de presidentes do Brasil (Bolsonaro, Lula e Fernando Henrique Cardoso).

As siglas dos partidos (PSL, PT e PSDB) e as palavras direita, esquerda e centro foram relacionadas corretamente às figuras abstratas, sem treino específico, o que os autores tomam como sugestão de que já faziam parte, antes do experimento, de uma classe com as fotografias dos presidentes; e, depois de aprender as relações entre fotografias e figuras abstratas, as avaliações das figuras, feitas numa escala de adjetivos opostos, ficaram mais próximas das da fotografia correspondente. Em outro estudo dos mesmos autores, com dois experimentos de cinco participantes cada, o aumento da semelhança entre as avaliações (100 menos a diferença absoluta entre elas) foi, em faces neutras de pessoas desconhecidas, a metade do obtido com figuras abstratas, o que atribuem, provavelmente, à carga afetiva prévia das faces (Pereira e Moreira, 2021a). Os autores dizem não conhecer estudo de atitudes com estímulos de significado social ligados a posicionamentos políticos.

Há uma ressalva de resultado, que o Capítulo 4 já deu. Haydu et al. (2015) ensinaram, por pareamento ao modelo, 28 homens de torcidas organizadas de três clubes brasileiros a relacionar os emblemas dos clubes a pinturas abstratas e essas pinturas às palavras bom, ruim e regular. Os testes avaliavam se emergiam, sem ensino direto, relações entre o emblema do clube apoiado e a palavra ruim, entre o do principal rival e a palavra bom e entre o do rival menor e a palavra regular, de modo que as palavras conflitavam com a história de cada torcida. Os participantes formaram as relações ensinadas (de linha de base) e as de simetria, mas não as de transitividade e de equivalência entre emblemas e palavras, e os erros mais frequentes foram escolher o que a história prescrevia: bom para o emblema do clube apoiado e ruim para o do principal rival. O que o estudo mostra, para este capítulo, é que a classe pretendida nem sempre se forma quando contraria a história do participante. Não se generaliza o achado à filiação partidária, e tampouco o de Pereira e Moreira.

A hipótese do capítulo é a seguinte. Em qualquer eleição em que a escolha exija reconhecer ou produzir um nome, uma sigla, uma imagem ou um número de candidato, o ato supõe uma discriminação condicional entre esses estímulos, e a análise de classes de equivalência trata de estímulos arbitrariamente relacionados por história, que se tornam substituíveis, com a função de um chegando aos outros sem contato direto (seção 4.3). Se nome, imagem, sigla e número do candidato fazem parte de uma mesma classe para o eleitor, a função avaliativa de qualquer um deles pode chegar aos demais, e o voto pode estar menos sob controle do que o candidato disse (Capítulos 3 e 5) do que das relações que o eleitor aprendeu entre os estímulos que o designam. Não há fonte lida sobre o procedimento de votação adotado no Brasil, e por isso ele não é descrito: a hipótese vale para qualquer procedimento que peça ao eleitor um desses estímulos.

O que testaria a hipótese é um estudo que medisse, para candidatos de campos políticos diferentes e com o mesmo procedimento para todos, as relações entre nome, imagem, sigla e número (por pareamento ao modelo ou por tarefa de nomeação), a função avaliativa de cada um (por diferencial semântico, como no Capítulo 4) e a transferência dessa função depois de uma história de pareamento. A regra de simetria vale: mesmo procedimento, mesmos critérios, para todos os campos.

■ DADO	Siglas de partidos e as palavras direita, esquerda e centro foram relacionadas corretamente a figuras abstratas, sem treino específico, em cinco estudantes (Pereira e Moreira, 2021b). Em 28 torcedores, relações de linha de base e de simetria se formaram, mas não transitividade e equivalência entre emblemas e palavras (Haydu et al., 2015).
■ INTERPRETAÇÃO	A transferência de função avaliativa entre fotografias, siglas e figuras é possível em laboratório e depende da história prévia com os estímulos; onde o que se ensina contraria a história do participante, a classe pretendida não emergiu (Haydu et al., 2015).
■ HIPÓTESE	Que o número do candidato, como o nome e a sigla, entre numa classe com a fotografia e com o rótulo direita-esquerda, e que isso mede o voto, é hipótese aqui proposta. Nenhuma fonte verificada manipulou números de candidatos nem mediu voto em eleitores de equivalência.

7.7 Interpretações rivais e limites

A leitura do capítulo, de que o voto é comportamento cuja manutenção é social e cuja relação com o que o eleitor diz depende de contingências próprias de cada um, tem rivais, e a Tabela 7.2 as reúne com o que as separaria dela.

Os limites do capítulo são os das fontes. Nenhuma mede o voto de eleitores brasileiros: o estudo de voto em termos operantes é um levantamento holandês de 1986, de que o próprio autor ressalva a causalidade, e o experimento é de uma plataforma, num país, numa eleição. A literatura de atitude e comportamento mede atitude, e não voto, e LaPiere e Wicker são conhecidos aqui só pelo relato de Lloyd. A classificação do voto como comportamento verbal é aqui proposta, e vale como análise, e não como achado. E nenhuma fonte compara campos políticos, de modo que o capítulo não diz nada sobre diferenças entre eleitores de campos diferentes.

TABELA 7.2*Leituras rivais do voto e do que o eleitor diz dele: o que as sustentaria e o que as separaria*

Leitura rival	O que a sustentaria	O que a separaria da leitura do capítulo
O voto é comportamento não verbal, sem mediação social; a declaração é outra resposta e a relação entre elas não é pergunta sobre o voto	Voto que não varia quando mudam as consequências sociais, nem depois de declarações	A mensagem social de Bond et al. mudou o voto validado; o modelo de Visser explica 65% dos votos
Ambiente compartilhado: pais e parceiro votam igual porque dividem informação, bairro e renda, e a consistência decorre disso, e não de reforço (Visser, 1996, sobre a causalidade)	Consistência igual entre eleitores que não são aprovados nem desaprovados por pais ou parceiro	Delineamento que mantenha o ambiente e varie a aprovação, ou que siga o eleitor no tempo
Atitude como disposição: declaração e voto são efeitos da mesma causa, e as inconsistências vêm de medida ruim (descrito por Street, 1994)	Correlações altas entre declaração e voto validado com instrumentos melhores	Dizer e fazer variando por razões diferentes (Risley e Hart, 1968; DeFleur e Westie, 1958)
Lembrete e informação: a mensagem social só lembrou de votar	Efeito igual com mensagem informativa	Bond et al.: a informativa teve o mesmo comparecimento do controle; mas o experimento não separa modelo, audiência e regra
Relato impreciso: a consistência de Visser vem de memória e de desejabilidade, e não de comportamento	Relato de votos passados que se aproxima do do voto presente	Voto validado em registro público, como em Bond et al.

Síntese

Votar é comportamento cuja manutenção é social, e por isso cabe a ele a análise do comportamento verbal, embora o ouvinte do ato seja um sistema, e não uma pessoa presente (7.1). Um mesmo voto pode estar sob controle de tato, de intraverbal, de mando ou de regra, e a forma não decide (7.2). O sigilo retira do ato a audiência identificada e deixa a declaração, que a tem; o que a pesquisa eleitoral observa é a declaração, sob a audiência da entrevista, e não o voto (7.3). A relação entre dizer e fazer é, desde LaPiere, baixa e variável, e a análise do comportamento a lê como a de duas classes de respostas com contingências próprias; a correspondência é efeito de uma contingência que a reforça, e reforçar o dizer sozinho não produz o fazer (7.4). Há dado de associação e dado de manipulação de que o que a comunidade faz muda a probabilidade de votar, mas o desenho não separa modelo, audiência e regra (7.5). A hipótese de que nome, imagem, sigla e número do candidato formem classe para o eleitor é aqui proposta, e nenhuma fonte a testa (7.6).

O Capítulo 10 (seção 10.1) trata por inteiro das audiências negativa e não punitiva de que a seção 7.3 se serve, e o 12 retoma a correspondência dizer-fazer ao avaliar intervenções cujo desfecho é o que o participante diz.

Referências

- Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415), 295–298. <https://doi.org/10.1038/nature11421>
- DeFleur, M. L., & Westie, F. R. (1958). Verbal attitudes and overt acts: An experiment on the salience of attitudes. *American Sociological Review*, 23(6), 667–673. <https://doi.org/10.2307/2089055>
- Guerin, B. (1994). Attitudes and beliefs as verbal behavior. *The Behavior Analyst*, 17(1), 155–163. <https://doi.org/10.1007/bf03392661>
- Haydu, V. B., Camargo, J., & Bayer, H. (2015). Effects of preexperimental history on the formation of stimulus equivalence classes: A study with supporters of Brazilian soccer clubs. *Psychology & Neuroscience*, 8(3), 385–396. <https://doi.org/10.1037/h0101276>
- Linn, L. S. (1965). Verbal attitudes and overt behavior: A study of racial discrimination. *Social Forces*, 43(3), 353–364. <https://doi.org/10.2307/2574765>
- Lloyd, K. E. (1994). Do as I say, not as I do. *The Behavior Analyst*, 17(1), 131–139. <https://doi.org/10.1007/bf03392658>
- Pereira, P. B. G., & Moreira, M. B. (2021a). *Atitudes em relação a presidentes do Brasil: Formação de classes de equivalência e transferência de função* (1ª ed.). Instituto Walden4. https://www.walden4.com.br/livros/pdfs/pereira_moreira_2021_presidentes.pdf
- Pereira, P. B. G., & Moreira, M. B. (2021b). Transferência de atitudes em relação a presidentes da república. *Acta Comportamentalia*, 29(4), 137–154. <https://doi.org/10.32870/ac.v29i4.80318>
- Regan, D. T., & Fazio, R. (1977). On the consistency between attitudes and behavior: Look to the method of attitude formation. *Journal of Experimental Social Psychology*, 13(1), 28–45. [https://doi.org/10.1016/0022-1031\(77\)90011-7](https://doi.org/10.1016/0022-1031(77)90011-7)
- Risley, T. R., & Hart, B. (1968). Developing correspondence between the nonverbal and verbal behavior of preschool children. *Journal of Applied Behavior Analysis*, 1(4), 267–281. <https://doi.org/10.1901/jaba.1968.1-267>
- Skinner, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts.
- Street, W. R. (1994). Attitude-behavior congruity, mindfulness, and self-focused attention: A behavior-analytic reconstruction. *The Behavior Analyst*, 17(1), 145–153. <https://doi.org/10.1007/bf03392660>
- Visser, M. (1996). Voting: A behavioral analysis. *Behavior and Social Issues*, 6(1), 23–34. <https://doi.org/10.5210/bsi.v6i1.278>



PARTE III

A ecologia da informação

-
- 8** Economia da atenção: o algoritmo como programador de contingências
 - 9** O ciclo produtor-audiência
 - 10** A janela de Overton: o que se pode dizer sem ser punido
 - 11** Infodemiologia: como a informação se espalha e o que se mede

8

Economia da atenção: o algoritmo como programador de contingências

O Upworthy, um site de notícias virais, testava manchetes em experimentos, e um deles serve de ponto de partida, mas será melhor abordado à frente. Uma mesma notícia foi apresentada a leitores com quatro manchetes diferentes, e as taxas de cliques foram de 0,94%, 0,79%, 0,60% e 0,58% (Robertson et al., 2023, Tabela 2). Com a mesma matéria e o mesmo veículo, a manchete mais clicada recebeu cerca de 1,6 vez os cliques da menos clicada (cálculo próprio). O site fazia isso em série: o conjunto de dados reúne 22.743 testes randomizados, com cerca de 105 mil variações de manchete.

Cada teste é um arranjo experimental limpo: a manchete varia, a notícia e o público são os mesmos, e o que se mede é a atenção que cada variação obtém, contada em cliques. O que acontece com o que se diz quando é assim que a atenção se distribui? O capítulo trata da economia em que essa distribuição ocorre. A seção 8.1 apresenta a atenção como reforçador; a 8.2, as propriedades do conteúdo que os dados associam a mais cliques, compartilhamentos e reações; a 8.3, a regra com que a plataforma decide quantos veem cada emissão; a 8.4, a hipótese de que essa entrega funcione como reforçamento intermitente; e a 8.5, os experimentos em que a regra foi manipulada.

O capítulo é cauteloso em dois pontos que a conversa corrente dá por resolvidos. O primeiro é que o feed seja um esquema de razão variável: nenhuma fonte verificada mede esquema algum, e a seção 8.4 diz o que a literatura dos esquemas descreve e o que faltaria para a comparação. O segundo é que o algoritmo premie a raiva ou favoreça um campo político: o que os dados mostram é mais estreito, e a seção 8.3 o descreve nos termos dos autores. Em todas as seções, dado, interpretação e hipótese ficam separados.

8.1 Atenção como reforçador

Skinner (1957) trata o comportamento verbal como comportamento mantido por consequências mediadas por ouvintes, e a atenção do ouvinte aparece em passagens dispersas do livro, a mais explícita na p. 199. O quadro reúne o que ele diz e o que este capítulo toma dali.

CONCEITO EM USO

Atenção como reforçador condicionado generalizado (Skinner, 1957)

O Capítulo 3 (seção 3.3) apresentou o reforçador condicionado generalizado: o evento que costuma preceder muitos reforçadores diferentes e que, por isso, pode reforçar o comportamento sob todas as condições apropriadas de privação e de estimulação aversiva, como a aprovação (Skinner, 1957). A atenção do ouvinte tem a mesma estrutura: é a consequência comum de mandos generalizados como *Como?* e *Onde!*, e, como o comportamento seguinte do ouvinte pode ser relevante para muitos estados de privação, essas respostas ficam sob controle amplo. Skinner a dá como exemplo de variável que fortalece o comportamento verbal sem respeito à forma: qualquer comportamento verbal que evoque atenção é reforçado, à parte outras ações específicas do ouvinte. O reforço social generalizado de Guerin (1994), que o Capítulo 3 também apresentou, é conceito vizinho: atitudes são enunciados mantidos por reforçamento social generalizado de uma comunidade verbal.

Dois pontos do quadro importam para o resto do capítulo. O primeiro é a razão de a atenção ser reforçador de amplo alcance: ela é condição para que o ouvinte faça qualquer outra coisa por quem fala, de modo que serve a muitos estados de privação ao mesmo tempo. O segundo é a cláusula “sem respeito à forma”. O que se fortalece é qualquer resposta que evoque atenção, e a atenção, como consequência, não distingue por si o verdadeiro do falso nem o que informa do que irrita. Esta última frase é uma leitura aqui proposta a partir de Skinner, que não trata de desinformação nem de plataformas. É ela, porém, que dá sentido a perguntar, nas seções seguintes, que propriedades do conteúdo atraem atenção e quem controla a sua distribuição.

Numa plataforma, a atenção do ouvinte não é observável por quem fala. O produtor vê contagens: número de seguidores, curtidas, compartilhamentos, comentários, cliques. A medida mais fina entre as fontes verificadas é a “impressão de permanência” do Twitter, o registro de que ao menos 50% da área de uma publicação ficou visível por ao menos 500 milissegundos, que Huszár et al. (2022) chamam de a melhor aproximação disponível para saber se o usuário foi exposto ao conteúdo. É medida de exposição, e não de atenção em sentido funcional; o resto do que as fontes trazem são reações explícitas. Este

capítulo chama de atenção medida a contagem de reações que a plataforma devolve. A distância entre ela e a atenção que reforça, a do ouvinte, é premissa de leitura, e não achado.

Há um estudo que pergunta diretamente quanto vale a atenção entre as razões para usar redes sociais. Gunnarsson et al. (2026) aplicaram a 328 universitários de uma universidade canadense (idade média de 22,7 anos; 68% mulheres) uma tarefa hipotética de compra de redes sociais. Os participantes imaginavam que só poderiam usar redes comprando o tempo, vendido por minuto, e diziam quantos minutos comprariam por dia a cada um de 15 preços, de 0,01 a 1.120 dólares por minuto, em quatro condições de função: usar as redes unicamente para escapar do tédio e de sentimentos desagradáveis (escape), para buscar informação (tangível), para obter engajamento com o próprio conteúdo, como curtidas, compartilhamentos e comentários (atenção), ou para se sentir bem (automática). A função afetou o consumo a custo mínimo (Q_0 , o que se compraria se o tempo custasse zero), mas não a sensibilidade ao preço (α , a rapidez com que a compra cai quando o preço sobe), e, na formulação do resumo dos autores, escape, informação ou reforço positivo automático geraram mais demanda que buscar curtidas. Na condição de atenção, 57,6% das trajetórias individuais falharam no critério de sistematicidade ΔQ , um dos indicadores de qualidade dos dados de demanda (13,4%, 20,7% e 11,6% nas demais condições), em sua maioria por consumo zero, e 51,5% dos participantes registraram apenas valores zero (7,0% na condição tangível). Os autores leem isso como baixa demanda por usar redes sociais para obter atenção, e não como desatenção à tarefa, mas ressaltam que a tarefa é hipotética.

O que o estudo diz ao capítulo é uma cautela. Se a atenção fosse, para quem usa redes, o reforçador dominante, esperar-se-ia alta demanda justamente na condição que a nomeia, e não foi o que apareceu, numa amostra de estudantes e numa tarefa que isolou uma função, embora muitas pessoas usem redes por funções múltiplas e sobrepostas (limitação declarada pelos autores). Há ainda uma diferença de construto: o estudo mede o valor, para o usuário, de usar a rede para obter atenção, e não a eficácia da atenção como consequência de uma publicação específica, que é a pergunta do capítulo. As duas coisas podem divergir.

■ DADO	Skinner (1957) descreve a atenção do ouvinte como consequência comum de mandos generalizados e como variável que fortalece o comportamento verbal sem respeito à forma; numa tarefa hipotética com universitários canadenses, a demanda por usar redes para obter atenção foi menor que a por escape, informação ou reforço positivo automático (Gunnarsson et al., 2026).
■ INTERPRETAÇÃO	A atenção do ouvinte é reforçador generalizado de amplo alcance para o comportamento verbal. Numa plataforma, o que chega ao produtor é uma contagem de reações, que serve de medida aproximada dessa atenção.
■ HIPÓTESE	A contagem de reações é o reforçador que mantém o que o produtor publica e é dominante em relação a outros. Os estudos do Capítulo 9 dão apoio parcial à primeira metade (curtidas e frequência de postar; retorno social e indignação), com efeitos pequenos e populações específicas; a segunda metade não foi mostrada, e o resultado de Gunnarsson et al. (2026) pede cautela.

8.2 O que os dados mostram sobre conteúdo emocional, moralizado e negativo

Três estudos verificados ligam propriedades do conteúdo à reação da audiência. Medem respostas diferentes (clique, compartilhar, publicar), e a diferença é parte do que há a dizer.

Robertson et al. (2023) analisaram os testes do Upworthy, em que cada visitante do site via ao acaso uma de várias variações de manchete de uma mesma notícia, e o clique para ler a matéria era registrado (visitantes entre 2013 e 2015). O conjunto analisado reúne 22.743 testes, dos quais 12.448 passaram pelo filtro pré-registrado, que descarta, por exemplo, os testes com uma única manchete (muitos testavam imagens de chamada, e não manchetes). Esses testes compararam em média 4,31 variações cada um, somando 53.699 manchetes, mais de 205 milhões de impressões e 2.778.124 cliques; a taxa média foi de 1,39%. As palavras das manchetes foram classificadas, uma a uma, por um dicionário de linguagem, o LIWC, com 620 palavras positivas e 744 negativas. Para uma manchete de tamanho médio, cada palavra negativa a mais aumentou a taxa de cliques em 2,3%, e a presença de palavras positivas a diminuiu em cerca de 1,0%. Os autores destacam que as palavras positivas eram um pouco mais prevalentes que as negativas, 2,83% contra 2,62% das palavras. O efeito das palavras negativas foi mais forte nas notícias de Governo e Econo-

mia. Entre as emoções discretas, a tristeza aumentou a probabilidade de clique e a alegria a diminuiu; a raiva não teve efeito estatisticamente significativo, e o medo teve efeito em sentido oposto ao pré-registrado.

O desenho tem limites que os próprios autores declaram. O Upworthy usava manchetes de caça-clique e difere das fontes tradicionais de notícia; as conclusões são do nível das notícias, e não de indivíduos; e nada dizem da experiência subjetiva dos leitores. As manchetes recebiam menos cliques quando publicadas mais tarde na história do site, o que os autores leem como sugestão de que o caça-clique era mais eficaz quando novo para os leitores. Além disso, o que o estudo mede é consumo, que eles descrevem como comportamento privado, ao contrário do compartilhamento, que é público e influenciado por fatores sociais, como sinalizar identidade de grupo e manter reputação; isso pode explicar que medo e raiva pesem mais na decisão de compartilhar. A explicação é dos autores e não foi testada nos dados do artigo.

Os outros dois estudos aparecem também no Capítulo 9 e entram aqui só para a comparação. Rathje et al. (2021) contaram, em 2.730.215 publicações de veículos de notícia e de congressistas dos Estados Unidos no Facebook e no Twitter, as palavras que se referiam ao campo político oposto (o exogrupo) e ao próprio campo (o endogrupo), e relacionaram essas contagens ao número de compartilhamentos. As publicações que se referiam ao exogrupo foram compartilhadas cerca de duas vezes mais, e cada termo sobre o exogrupo se associou, em média, a um aumento de cerca de 67% (que os autores descrevem como aumento da chance de compartilhar), efeito 4,8 vezes o da linguagem de afeto negativo e 6,7 vezes o da linguagem moral-emocional (seção 9.3). Brady et al. (2021) estudaram a expressão de indignação do próprio produtor. Em 12,7 milhões de tweets de 7.331 usuários, um aumento de 100% no retorno social (curtidas e compartilhamentos) à indignação do dia anterior associou-se a um aumento esperado de 2 a 3% na indignação do dia seguinte; em dois experimentos de 120 participantes, que escolhiam tweets para postar numa rede simulada, a escolha de tweets de indignação cresceu com o retorno e acompanhou a norma exibida no feed (seção 9.5).

A comparação ensina mais pelo que impede do que pelo que permite. “Conteúdo negativo é mais reforçado” não é uma proposição que descreve um processo único; são ao menos três: manchetes com palavras negativas são mais clicadas (Robertson et al.), publicações que mencionam o exogrupo são mais compartilhadas (Rathje et al.) e a expressão de indignação acompanha o retorno recebido (Brady et al.). São respostas distintas, de pessoas em posições distintas do ciclo que será visto no Capítulo 9, e nada obriga que a mesma consequência as mantenha. O padrão das emoções discretas mostra a diferença: a raiva, que a ideia corrente põe como motor da viralidade, não teve efeito signifi-

ficativo nos cliques, ao passo que a linguagem do exogrupo previu reações “angry” nas publicações do Facebook (Rathje et al.). A explicação dos autores de que consumir é privado e compartilhar é público é uma hipótese funcional, e o capítulo a toma com esse nome.

TABELA 8.1

Três respostas ao conteúdo emocional e moralizado: o que cada estudo mediu e o que não permite dizer

Estudo	Resposta e material	Propriedade do conteúdo	O que observaram	Ressalva
Robertson et al. (2023)	Clicar na manchete (consumo, comportamento privado). 12.448 testes randomizados do Upworthy; 53.699 manchetes; mais de 205 milhões de impressões	Palavras negativas e positivas e emoções discretas, por dicionário	Cada palavra negativa: +2,3% de cliques (manchete de tamanho médio); positivas: cerca de -1,0%; tristeza aumentou, alegria diminuiu; raiva sem efeito significativo	Veículo de caça-clique; nível da notícia, não do indivíduo; nada sobre a experiência do leitor
Rathje et al. (2021)	Compartilhar e reagir (comportamento público). 2.730.215 publicações de veículos de notícia e de congressistas dos EUA, no Facebook e no Twitter	Referência ao exogrupo político; linguagem de afeto negativo e moral-emocional	Exogrupo: cerca de duas vezes mais compartilhamento; +67% por termo, 4,8 e 6,7 vezes o efeito do afeto negativo e do moral-emocional; reação “angry” mais popular em sete de oito conjuntos	Correlacional; período anterior à eleição de 2020; não está claro como as escolhas algorítmicas contribuem
Brady et al. (2021)	Publicar indignação (comportamento do produtor). 12,7 milhões de tweets de 7.331 usuários; dois experimentos de 120 participantes	Indignação moral expressa, medida por classificador	+2 a 3% de indignação no dia seguinte para cada +100% de retorno; nos experimentos, a escolha de indignação cresceu com o retorno e acompanhou a norma	Estudos observacionais não permitem inferência causal; usuários identificados por episódios de indignação pública; mede texto, não emoção

Nota. Os números vêm dos artigos. As páginas de Robertson et al. (2023) são as de *Nature Human Behaviour*; Rathje et al. (2021) e Brady et al. (2021) não têm paginação de revista. A coluna “Propriedade do conteúdo” descreve forma (palavras, referência a um grupo, expressão classificada); a função que a forma cumpre para quem clica, compartilha ou publica não foi medida em nenhum dos três.

Há ainda a diferença entre forma e função. As propriedades medidas são de forma: palavras de um dicionário, a referência a um grupo, uma expressão de indignação classificada por modelo semântico. A análise operante pergunta pela função, isto é, o que faz de uma palavra negativa na manchete uma ocasião para clicar, ou de uma menção ao exogrupo uma ocasião para repassar. Duas respostas aparecem nos autores, sinalizar identidade de grupo e manter reputação (Robertson et al.) e satisfazer motivos de identidade social (Rathje et al.), e ambas são funções propostas que nenhum estudo manipula. Sem a função, “conteúdo negativo” é topografia, e o tratamento por função apresentado no Capítulo 3, mostra como muitas vezes a forma apresenta funções distintas ou está sob controle múltiplo.

■ DADO	Em testes randomizados com mais de 2,7 milhões de cliques, palavras negativas aumentaram a taxa de cliques e palavras positivas a diminuíram (Robertson et al., 2023); publicações sobre o exogrupo político são compartilhadas cerca de duas vezes mais que as sobre o endogrupo (Rathje et al., 2021); a indignação do dia seguinte acompanha o retorno recebido pela do dia anterior, com efeito pequeno (Brady et al., 2021).
■ INTERPRETAÇÃO	Propriedades de forma do conteúdo (palavras, referência ao exogrupo) associam-se a mais cliques, compartilhamentos e reações, e a associação difere conforme a resposta medida, o que sugere operantes distintos com consequências distintas.
■ HIPÓTESE	O que faz do conteúdo moralizado uma ocasião para repassar é a função que ele cumpre para quem repassa, por exemplo sinalizar identidade de grupo ou manter reputação (leitura dos autores). Nenhum dos estudos manipula a função.

8.3 O algoritmo como programador: quanta audiência cada emissão recebe

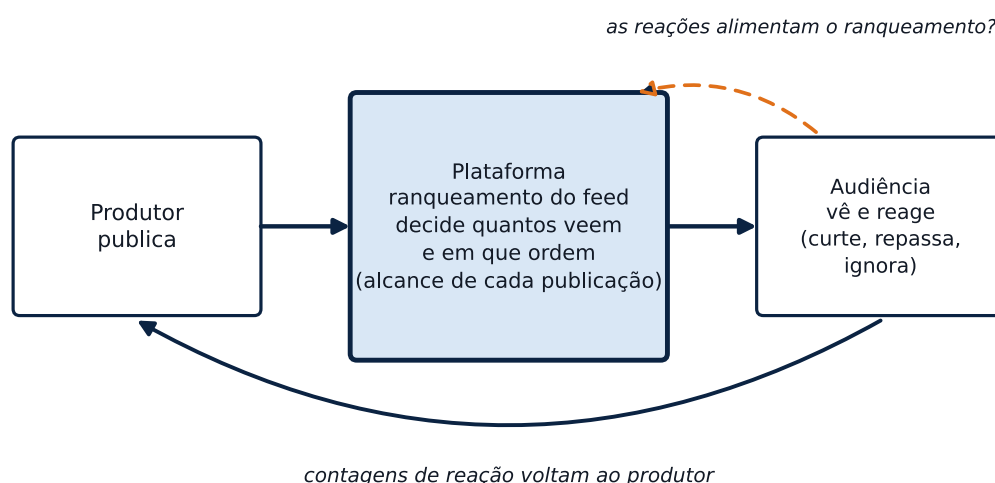
Nas seções anteriores, o produtor publica e a audiência reage; falta o termo que decide quantos veem cada emissão e em que ordem, e é dele que trata esta seção. Numa plataforma com feed ranqueado, esse termo é um conjunto de modelos de aprendizado de máquina; em termos operantes, é a regra que fixa, para cada emissão, o tamanho da audiência que poderá reagir a ela. Skinner (1957) já notava que os efeitos enfraquecedores do reforçamento intermitente ou atrasado são, no comportamento verbal, em parte compensados pelo fato de que os efeitos do comportamento verbal podem ser multiplicados, ao expor

muitos ouvidos às mesmas ondas sonoras ou muitos olhos à mesma página. A plataforma controla essa multiplicação: o ranqueamento decide quantos olhos a acessarão e reagirão a ela.

A Figura 8.1 é um modelo, e a nota diz onde há dado. As setas contínuas entre produtor, plataforma e audiência têm apoio na auditoria de Huszár et al. (2022), descrita a seguir, e a que leva as contagens de volta ao produtor, nos estudos do Capítulo 9. A tracejada, em que as reações alimentam o ranqueamento, é hipótese. Brady et al. (2021) afirmam que algoritmos de feed, ao determinar quantos usuários veem uma publicação, afetam o retorno social e podem influenciar comportamentos morais explorando tendências de aprendizagem por reforço; é interpretação dos autores, e o estudo não mede o algoritmo.

FIGURA 8.1

O algoritmo como programador das contingências do feed



Linha contínua: relação com dados (alcance sob ranqueamento e sob ordem cronológica, Huszár et al., 2022; sensibilidade de quem publica ao retorno, Lindström et al., 2021). Tracejada: hipótese; o que o ranqueamento otimiza não é conhecido nas fontes do capítulo.

Nota. Esquema original do autor, sem dados. Linha contínua: relação com dados (alcance sob ranqueamento e sob ordem cronológica, Huszár et al., 2022; sensibilidade de quem publica ao retorno, Lindström et al., 2021). Tracejada: hipótese; o que o ranqueamento usa como entrada não é conhecido nas fontes do capítulo.

Huszár et al. (2022) são a fonte verificada que compara o ranqueamento personalizado com a ordem cronológica do algoritmo sobre a audiência. Em 2016 o Twitter introduziu algoritmos de aprendizado de máquina para ordenar a Home timeline por um modelo de relevância personalizada; antes, ela mostrava, em ordem cronológica inversa, os tweets das contas seguidas. Ao introdu-

zir a personalização, excluiu dela um grupo de controle de 1% dos usuários do mundo, que seguiu com a linha do tempo cronológica; o tratamento era uma amostra de 4% das demais contas. Em 5 de junho de 2020 o experimento incluía 58.087.969 identificadores de conta, 5% das contas do mundo, 20% no controle e 80% no tratamento. A amplificação é a razão entre o alcance de um conjunto de tweets no grupo de tratamento e no de controle: 0% equivale a alcance igual, e valores acima de 200% indicam audiência três vezes a da linha do tempo cronológica. O alcance foi contado por impressões de permanência (seção 8.1).

Todos os partidos analisados tiveram amplificação positiva pela personalização, em alguns casos acima de 200%. Comparando o maior partido de direita ou centro-direita e o maior de esquerda ou centro-esquerda de cada país, com exceção da Alemanha, o primeiro teve amplificação significativamente maior; o efeito foi mais forte no Canadá (liberais 43%, conservadores 167%) e no Reino Unido (trabalhistas 112%, conservadores 176%). Os sete países são Estados Unidos, Japão, Reino Unido, França, Espanha, Canadá e Alemanha; o Brasil não está entre eles.

O resultado muda de natureza quando a unidade é o político, e não o partido. No nível individual, a amplificação variou muito dentro de cada partido, de cerca de 400% a abaixo de 0%, e um teste de permutação (que compara o resultado com o obtido ao embaralhar ao acaso os partidos entre os políticos) não detectou associação estatisticamente significativa entre o partido e a amplificação de cada político. Os autores não veem contradição: políticos diferentes podem alcançar audiências sobrepostas, de modo que correlações de grupo podem surgir mesmo sem correlação individual. Nos países com partidos de extrema esquerda ou de extrema direita bem representados, a amplificação desses partidos foi em geral menor que a dos moderados, e a análise não sustenta a hipótese de que a personalização amplifique ideologias extremas mais que vozes moderadas; os autores ressaltam que forte viés partidário nas notícias se associou a maior amplificação, o que não implica promoção de ideologia extrema. Também não houve padrão consistente entre governo e oposição: no Reino Unido a amplificação favoreceu os conservadores no governo e, no Canadá, os conservadores da oposição.

Para notícias, a análise usou 100.575.284 tweets que apontavam para 6.258.032 artigos, classificados pelo AllSides, uma base que rotula a posição política de cada veículo numa escala de cinco pontos, da esquerda partidária à direita partidária: fontes mais partidárias foram mais amplificadas que as de centro, e a direita partidária um pouco mais que a esquerda partidária. Com a base Ad

Fontes, outra base independente de rótulos, a esquerda partidária teve amplificação baixa (10,5%), e os autores dizem que o resultado depende da base de rótulos políticos.

As limitações vêm declaradas. O grupo de controle não é isolado de efeitos indiretos, porque quem está nele vê conteúdo compartilhado por quem está no tratamento; por isso o experimento não cumpre a pressuposição conhecida como SUTVA (de que o resultado de uma conta não dependa do tratamento das outras), o que impede estimativas não enviesadas de efeitos causais, e as estatísticas provavelmente os subestimam. Além disso, o tratamento, isto é, a experiência de ranqueamento, mudou ao longo do tempo. Quanto ao mecanismo, os autores dizem que entendê-lo exige estudo adicional e citam, como explicação possível, que partidos diferentes podem seguir estratégias diferentes no Twitter. O grupo de controle, por fim, não foi criado para pesquisa, e sim para melhorar o algoritmo e dar uma linha de base ao negócio; o trabalho foi revisado pelas equipes jurídica e de privacidade do Twitter, e não por comitê de ética.

Mosleh e Rand (2022), descritos no Capítulo 1, tocam o mesmo problema pelo lado da audiência. Construíram escores de falsidade para 816 elites públicas dos Estados Unidos a partir das checagens do PolitiFact: cada checagem recebe uma nota de veracidade (de 1, verdadeiro, a 0, “pants on fire”), e a média das notas da figura, subtraída de 1, dá o escore de falsidade. Construíram também, para cada usuário do Twitter, um escore de exposição baseado em quem ele segue. A exposição à desinformação das elites correlacionou-se negativamente com a qualidade das notícias que o próprio usuário compartilha e positivamente com a ideologia conservadora estimada. Os autores dizem que a abordagem se baseia na rede de contas seguidas e não capta exposição além delas, como por retuïtes ou por recomendações algorítmicas, e sugerem que o escore pode servir, em trabalhos futuros, para distinguir o papel das recomendações algorítmicas e o das preferências individuais. Os escores de falsidade vêm de um único verificador e dependem do que ele checkou, ponto de seleção que o Capítulo 2 tratou.

O que a auditoria mede é a diferença entre dois programadores, o ranqueamento personalizado e a ordem cronológica, e não o que o ranqueamento otimiza nem se as reações entram no cálculo. Para o argumento do capítulo isso basta para uma conclusão limitada: a plataforma é uma variável que altera o alcance de cada emissão, e a alteração não é uniforme entre emissores, de modo que a análise do produtor precisa de duas variáveis, o que a audiência faz com o que recebe e quanto recebe, e a segunda é fixada pelo programador. Que a diferença entre grupos reflita preferência da plataforma, estratégia dos

partidos ou comportamento da audiência é o que os dados não decidem; usa-se o mesmo critério para todos os campos: amplificação é dado de alcance, e não de viés nem de intenção.

■ DADO	Em seis de sete países, o maior partido de direita ou centro-direita teve amplificação pela personalização significativamente maior que o maior de esquerda ou centro-esquerda; na Alemanha a diferença não foi significativa; por político, não houve associação entre partido e amplificação; o resultado para notícias depende da base de rótulos (Huszár et al., 2022).
■ INTERPRETAÇÃO	A plataforma é uma variável que determina o alcance de cada emissão de modo desigual entre produtores, e o alcance entra no que o produtor pode receber (Skinner, 1957). A diferença entre os dois ordenamentos não revela o que o ranqueamento otimiza.
■ HIPÓTESE	Se as reações recebidas realimentam o ranqueamento, a plataforma fecha o laço da Figura 8.1: o que foi reagido passa a ser mostrado a mais gente, e o produtor recebe mais do mesmo. Nenhuma fonte do capítulo mostra o que o ranqueamento usa como entrada.

8.4 Reforçamento intermitente e engajamento

A analogia mais repetida sobre redes sociais é a da máquina caça-níqueis: o usuário puxa a alavanca, e o prêmio vem de vez em quando. Esta seção diz de onde vem o vocabulário da analogia, o que a literatura experimental descreve para cada esquema e o que as fontes verificadas sustentam sobre o feed, que é pouco.

CONCEITO EM USO

Reforçamento intermitente (Skinner, 1957)

Reforçamento intermitente é aquele em que só parte das respostas, ou das ocasiões, é seguida do reforçador. Para Skinner (1957), o reforçamento meramente intermitente, como o das jogadas casuais de dados, basta para manter uma resposta em força; ele remete, para isso, ao estudo experimental do comportamento não verbal. No comportamento verbal, a intermitência é a regra: o efeito depende da presença e da atividade de um organismo reforçador, o ouvinte, cujo comportamento não é inevitável nem em geral previsível. Os esquemas nomeados (razão e intervalo, fixos e variáveis), de que a analogia do caça-níqueis toma o nome, vêm da literatura experimental que o quadro seguinte resume.

O argumento de Skinner tem duas consequências para o capítulo. A primeira é que, para quem fala, a intermitência é condição normal, e não peculiaridade das redes: o ouvinte pode não estar, pode não reagir, e o efeito sobre quem fala é intermitente por construção. A segunda é que o que a plataforma muda não é a presença da intermitência, e sim a sua forma: quantos ouvintes podem reagir (seção 8.3) e a visibilidade da contagem para o produtor e para os outros ouvintes. Esta é a leitura aqui proposta.

CONCEITO EM USO

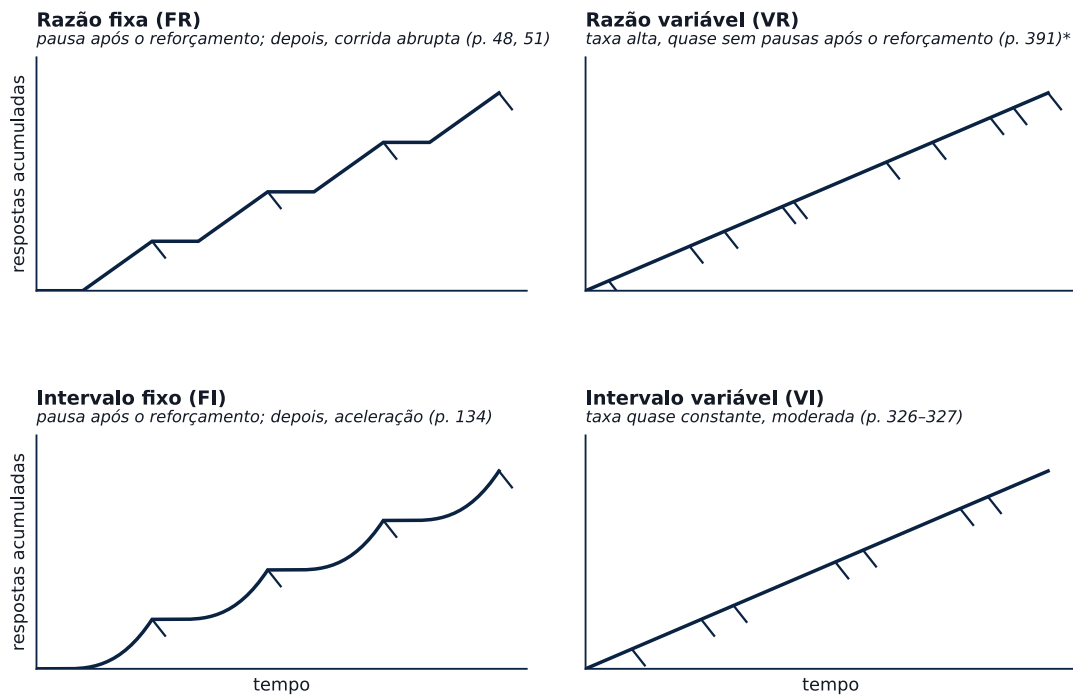
Esquemas de reforçamento (Ferster e Skinner, 1957)

Um esquema de reforçamento pode ser definido sem referência ao seu efeito sobre o comportamento: o reforçador pode depender do tempo transcorrido desde o reforçamento anterior ou do número de respostas emitidas desde então, e o esquema pode ser fixo ou variar, o que dá quatro esquemas básicos. Na razão fixa (FR), a resposta é reforçada ao completar um número fixo de respostas contadas desde o reforçamento anterior; na razão variável (VR), segundo uma série aleatória de razões com dada média; no intervalo fixo (FI), a primeira resposta depois de dado intervalo, medido desde o reforçamento anterior; no intervalo variável (VI), segundo uma série aleatória de intervalos com dada média. Os autores distinguem a função de moldar a forma da resposta e o controle de estímulos da função de manter um repertório ativo, e dizem que a maioria dos reforçamentos é intermitente. A maior parte dos experimentos usa pombos que bicam uma tecla para obter alimento, e alguns usam ratos que pressionam uma barra. Para o nível humano, dizem que a análise de esquemas se mostrou útil no estudo do comportamento psicótico e em técnicas educacionais em sala de aula, e que outras aplicações, em direito e penologia, religião, indústria e comércio, oferecem promessa considerável.

O que cada esquema gera, nos experimentos dos autores, é o termo de comparação de que a analogia precisa. Na razão fixa estabilizada depois de longa exposição, as pausas após o reforçamento são ocasionais e, quando ocorrem, a passagem à taxa terminal é abrupta (FR 60, depois de mais de 4.000 reforçamentos); com razões maiores as pausas crescem, até 10 minutos em FR 120. No intervalo fixo, o estado estável normal tem uma pausa após cada reforçamento, seguida de aceleração até uma taxa terminal em geral moderada. O intervalo variável é desenhado para produzir taxa constante, ao impedir que algum aspecto do comportamento da ave adquira propriedades discriminativas, e a maioria dos VI inclui intervalo zero ou próximo de zero, para evitar a pausa após o reforçamento; numa ave em VI 1 minuto, depois de 16 sessões, a taxa ficou aproximadamente constante, sem variação correlacionada a característica alguma do esquema. A razão variável, como a fixa, arranja reforçamento diferencial de taxas altas, e há poucos casos de reforçamento depois de pausa. A Figura 8.2 traça esses padrões num esquema.

FIGURA 8.2

Padrões de resposta que Ferster e Skinner (1957) descrevem para os quatro esquemas básicos: esquema conceitual



Esquema original, sem dados: inclinações e durações são ilustrativas, as escalas verticais são independentes (não compare alturas entre painéis) e as marcas oblíquas indicam reforçamentos. Traça o que Ferster e Skinner (1957) descrevem em texto, para pombos que bicam uma tecla para obter alimento.

** Com médias grandes, a razão variável gera variedade, com pausas longas e rajadas (p. 394-397).*

Nota. Esquema original do autor, sem dados, a partir das descrições em texto de Ferster e Skinner (1957), cujos registros são de pombos. O registro cumulativo, em que as respostas se somam ao longo do tempo, é o dado básico dos autores. Inclinações e durações são ilustrativas, e a escala vertical de cada painel é independente.

A razão variável, o esquema que a analogia invoca, é também o de que os autores registram maior variedade: produz performances diversas conforme a distribuição dos números de respostas exigidas, e a ave deixa de responder se a média passa de certo valor. Com média de 360 respostas, uma ave manteve numa sessão taxa sustentada de cerca de 7 respostas por segundo e, na seguinte, uma rajada logo após o reforçamento e depois pausas longas, uma de mais de 1 hora; os autores tratam as duas performances como os limites de um estado estável. A outra ave nunca sustentou taxa alta e respondia em rajadas de 10 a 75 respostas, separadas por pausas de 10 a 70 segundos.

Gunnarsson et al. (2026) escrevem, na discussão, que seus achados ilustram como as plataformas empregam esquemas de razão variável: como as curtidas não são garantidas, cada episódio de atenção bem-sucedido torna-se desproporcionalmente reforçador. É interpretação dos autores; o esquema não foi

manipulado nem medido, e não poderia ter sido numa tarefa hipotética de compra. O argumento também salta da intermitência para a razão variável: que as curtidas não sejam garantidas mostra que o reforçamento é intermitente, e chamá-lo de razão variável exige, além disso, que a entrega dependa do número de respostas e varie em torno de uma média (Ferster e Skinner, 1957). Os autores sugerem ainda que quem usa redes principalmente por reforçamento ligado à atenção pode ser particularmente vulnerável a padrões de uso problemático, a partir da interação entre a escala de uso de redes e a condição de função para o consumo a custo mínimo.

Dois estudos do Capítulo 9 são compatíveis com a hipótese sem testá-la. Lindström et al. (2021) mostram, em dados do Instagram e de fóruns, que a taxa de postar acompanha a taxa de curtidas segundo uma função hiperbólica, que sobe depressa e depois se achata (compatível com a lei quantitativa do efeito), melhor do que segundo uma relação linear. Ceylan et al. (2023) recompensaram com pontos, em 80 tentativas, a escolha de compartilhar ou não manchetes conforme um critério de veracidade e mostram que o padrão de compartilhar, instalado por esse retorno, persiste no teste em que o retorno é retirado. Em nenhum dos dois a intermitência foi a variável manipulada: a persistência sem retorno é compatível com uma história de reforçamento intermitente e com outras histórias, e o desenho não as separa. Lindström et al. tratam de quando se posta, e não do que se posta, e ressaltam que há muitos motivos para postar além da recompensa (seção 9.2).

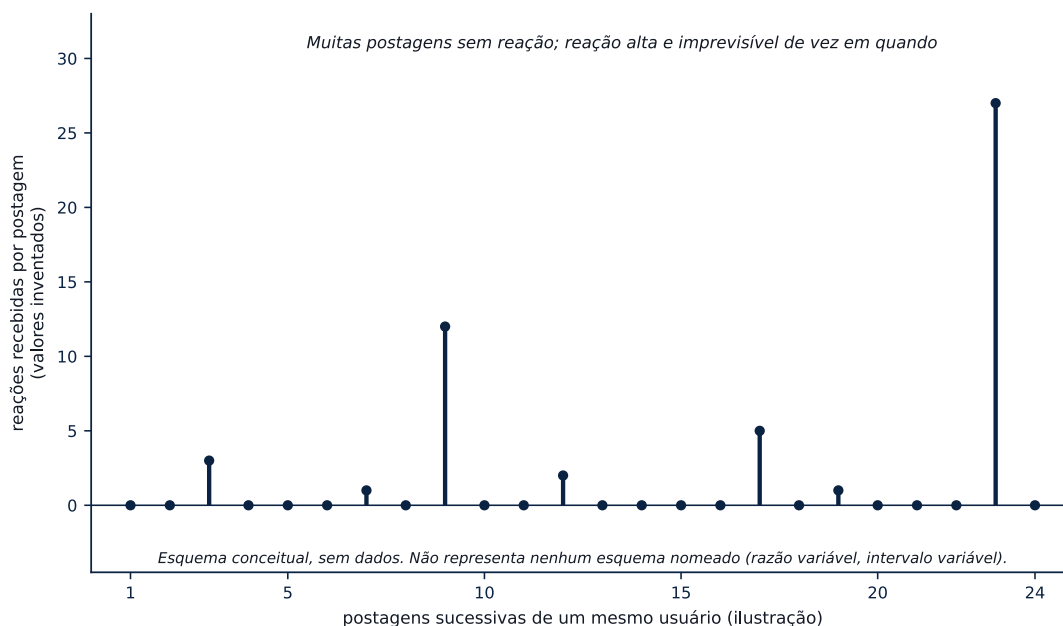
Chamar o feed de esquema de razão variável exige, no mínimo, três coisas. A primeira é a distribuição das reações por publicação, para um mesmo produtor, ao longo de muitas publicações: razão variável supõe uma série de razões com dada média, e não só uma contagem que oscila. A segunda é a relação entre o número de respostas, publicar ou conferir as reações, e a entrega do reforçador: na razão, o que conta é o número de respostas desde o reforçamento anterior, e no intervalo, o tempo. A terceira é o padrão de resposta que o esquema gera, comparado com o da Figura 8.2 e com uma linha de base, por exemplo o tempo até a publicação seguinte depois de uma muito reagida e depois de uma ignorada. Nenhuma fonte do capítulo traz a primeira nem a segunda, e a terceira existe, para o feed, só como pergunta. A Figura 8.3 serve para ver os pontos frouxos da analogia, com valores inventados.

Três pontos frouxos aparecem na figura. O primeiro é que a reação a uma publicação não é um reforçador discreto que vem ou não vem, como a ficha de uma máquina ou a operação do comedouro nos experimentos dos autores: é uma contagem, que cresce com o tempo e depende de quantos veem (seção 8.3). O segundo é que há ao menos dois operantes candidatos, publicar e conferir as reações, e é possível que estejam sob regras de entrega diferentes: o

reforçador de conferir está disponível desde que alguém reaja, e o de publicar depende da publicação. Os autores estudaram, em experimentos que descrevem como exploratórios, duas respostas com esquemas próprios e simultâneos, e a de razão dominou a de intervalo; que algo parecido ocorra entre publicar e conferir também é pergunta, e não resultado. O terceiro é o mais importante para esta análise: se o que determina quanto cada publicação recebe é, em parte, a sua forma, e não o acaso, a contingência é diferencial, pois seleciona formas e não apenas doseia a frequência, o que sugeriria pelo menos que esquemas diferentes estariam sobrepostos ou concorrendo entre si. Um esquema de reforçamento descreve o quanto e quando; a seleção de formas descreve o quê, e os autores distinguem a modelagem da forma da resposta da manutenção do repertório. A seleção de formas é a que importa para a pergunta deste capítulo, e a que as fontes do capítulo sustentam melhor, na seção 8.2. Tudo isto é interpretação.

FIGURA 8.3

Reações recebidas por postagens sucessivas de um mesmo usuário: esquema conceitual



Nota. Esquema original do autor. Os valores foram inventados para mostrar a ideia de reforçamento intermitente e não representam dado, nem um esquema nomeado (razão variável, intervalo variável).

<p>■ DADO</p>	<p>Para Skinner (1957), a intermitência é a regra do comportamento verbal e basta para manter uma resposta; Ferster e Skinner (1957) definem os quatro esquemas básicos e descrevem, em pombos, um padrão para cada um (pausa e corrida na razão fixa, pausa e aceleração no intervalo fixo, taxa quase constante no intervalo variável, taxa alta com pausas raras na razão variável, com variedade quando a média é grande); Gunnarsson et al. (2026) atribuem às plataformas esquemas de razão variável sem medi-los; a taxa de postar acompanha a de curtidas segundo uma função hiperbólica (Lindström et al., 2021).</p>
<p>■ INTERPRETAÇÃO</p>	<p>A intermitência é condição geral do comportamento verbal. Que a entrega de reações a quem publica seja intermitente é plausível e coerente com os dados, mas é premissa, e não resultado; e intermitência não é, por si, razão variável.</p>
<p>■ HIPÓTESE</p>	<p>A entrega segue um esquema nomeado (razão variável ou intervalo variável) e o padrão de uso decorre disso. Exigiria a distribuição de reações por publicação, a relação entre respostas e entrega e o padrão de resposta comparado ao da Figura 8.2; nenhuma fonte do capítulo traz os dois primeiros, e o terceiro não foi examinado para o feed.</p>

8.5 Experimentos com o próprio algoritmo: o que mudou e o que não mudou

Os experimentos em que a regra de entrega foi manipulada são a evidência mais direta do que a plataforma faz. As fontes verificadas trazem dois, o de Huszár et al. (2022), descrito na seção 8.3, e o de Nyhan et al. (2023), além do teste do publicador de Robertson et al. (2023), que manipula outro ponto da cadeia. A Tabela 8.2 os põe lado a lado.

O experimento de Nyhan et al. (2023) está descrito na seção 6.2, e aqui interessa o que manipulou. Para os participantes designados ao tratamento, todo o conteúdo de amigos, grupos e páginas que se previa compartilhar a inclinação política do participante foi rebaixado no feed, de 24 de setembro a 23 de dezembro de 2020; o do grupo de controle não foi alterado sistematicamente. Segundo os autores, o único objetivo da intervenção era reduzir a exposição a fontes de mesma inclinação, com a maior força de rebaixamento que um pré-teste indicou que não deixaria o feed quase vazio. A exposição caiu de 53,7% para 36,2%, e não houve efeito mensurável nas oito medidas de atitude pré-registradas; quem teve a exposição reduzida engajou-se mais, por exposição, com o conteúdo de mesma inclinação que ainda via.

TABELA 8.2

Três manipulações ao longo da cadeia produtor–plataforma–audiência: o que mudou e o que não mudou

Estudo	O que foi manipulado	O que mediram	O que mudou	O que não mudou ou não foi medido
Robertson et al. (2023)	Variações de manchete da mesma notícia, atribuídas ao acaso a leitores do Upworthy (manipulação do publicador, não da plataforma)	Taxa de cliques de cada variação	Palavras negativas aumentaram os cliques (2,3% por palavra, em manchete de tamanho médio); palavras positivas os diminuíram (cerca de 1,0%)	O que o publicador passou a produzir; o indivíduo; a experiência do leitor
Huszár et al. (2022)	Ordem do feed: ranqueamento personalizado (tratamento, 4% das contas) contra ordem cronológica (controle, 1%)	Alcance, em impressões de permanência, de tweets de legisladores de sete países e de links de notícias	Alcance maior com a personalização para todos os partidos; maior para o maior partido de direita ou centro-direita em seis dos sete países	Associação entre partido e amplificação, por político; mecanismo causal; efeitos indiretos sobre o controle; o que os produtores fizeram depois
Nyhan et al. (2023)	Rebaixamento, no feed do Facebook, do conteúdo de fontes de mesma inclinação (24/09 a 23/12/2020; n = 23.377)	Parcela de conteúdo de mesma inclinação e de fontes cruzadas; oito medidas de atitude pré-registradas	Exposição a fontes de mesma inclinação de 53,7% para 36,2%, e a fontes cruzadas de 20,7% para 27,9%; mais engajamento, por exposição, com o conteúdo congenial restante	Atitudes (efeitos de $\pm 0,12$ desvio-padrão ou mais descartados); efeitos de uso prévio e acumulados em anos; o produtor

Nota. Os números vêm dos artigos. As páginas de Robertson et al. (2023) são as de *Nature Human Behaviour* e as de Nyhan et al. (2023), as de *Nature*; Huszár et al. (2022) foi lido sem paginação impressa. “Mudou” e “não mudou” dizem respeito ao que cada estudo mediu.

A seção 6.1 deixou uma pergunta sobre a fronteira da bolha: quanto do que vem de fora da comunidade verbal chega ao falante. A regra de entrega é uma das variáveis que a decidem, e o experimento mostra que ela a move: com o rebaixamento do conteúdo congenial, a exposição a fontes cruzadas subiu de

20,7% para 27,9%, e a de fontes nem congeniais nem cruzadas, de 25,6% para 35,9%. O que o experimento não mede é o que a figura da seção 6.1 perguntava: se a correção dirigida ao que o falante disse passou a chegar a ele. Exposição a fontes cruzadas é parcela do feed, e não correção do enunciado.

O que os dois experimentos têm em comum é o que não medem. Ambos intervêm na entrega e medem o que o ouvinte recebe ou declara: o alcance de tweets de políticos, num caso; a exposição e as atitudes de usuários, no outro. Nenhum mede o que o produtor fez depois de ver o seu alcance mudar. O ranqueamento de Huszár et al. alterou o tamanho da audiência de cada emissão, e o rebaixamento de Nyhan et al. diminuiu o alcance do conteúdo de mesma inclinação; para o produtor, ambos são mudanças na contingência de alcance. A pergunta de que o capítulo partiu, se e como quem publica ajusta o que publica a essa regra, só se responderia com séries de publicações de produtores antes e depois de mudanças na regra de entrega. As plataformas têm esses registros; os estudos verificados não os trazem, o que limita acessar as possíveis funções atencionais diferenciais envolvidas.

Nada disso autoriza concluir que a regra de entrega não importa (Nyhan et al. não encontraram efeito em atitudes declaradas, depois de três meses e num único serviço) nem que ela explique o que se diz (Huszár et al. mediram alcance, e não o que se diz). O efeito nulo em atitudes tem, aliás, explicações que os próprios autores listam: o conteúdo político congenial é fração do que se vê, o Facebook é fração da dieta de informação, persuadir é difícil e o desenho não capta o uso prévio nem efeitos acumulados ao longo de anos.

<p>■ DADO</p>	<p>Ranqueamento personalizado e ordem cronológica diferem no alcance de partidos e de fontes de notícia (Huszár et al., 2022); reduzir em um terço a exposição a conteúdo de mesma inclinação não alterou oito medidas de atitude em três meses (Nyhan et al., 2023).</p>
<p>■ INTERPRETAÇÃO</p>	<p>A regra de entrega altera o alcance, e o alcance é variável de que depende o que o produtor pode receber. Nos dois experimentos a variável dependente é do ouvinte ou do alcance, e não do produtor.</p>
<p>■ HIPÓTESE</p>	<p>Produtores ajustam o que publicam a mudanças no alcance que a regra de entrega produz. Para testá-la, seria preciso medir a produção antes e depois de uma mudança de regra; nenhuma fonte verificada o faz.</p>

8.6 Interpretações rivais e limites

A leitura do capítulo, de que a atenção distribuída pelo algoritmo funciona como reforçador de quem publica, tem rivais. O quadro abaixo as reúne, uma a uma, e os limites do capítulo vêm depois dele.

Interpretações rivais para o mesmo conjunto de dados

O usuário escolhe, e o algoritmo reflete. O que a audiência reage e o que o algoritmo mostra se confundem: Mosleh e Rand (2022) dizem que seu score não capta a exposição por recomendação algorítmica e sugerem que ele sirva para separar o papel das recomendações e o das preferências; Nyhan et al. (2023) notam que as mudanças testadas no algoritmo não anulam a propensão dos usuários a buscar informação congenial; e Rathje et al. (2021) dizem não saber como as escolhas algorítmicas contribuem.

Atenção não é o único reforçador, nem o principal. Na tarefa de Gunnarsson et al. (2026), escape, informação e reforço positivo automático geraram mais demanda que buscar curtidas, e os autores lembram que muitas pessoas usam redes por funções múltiplas e sobrepostas.

Norma e identidade em lugar de reforço. O que se observa pode ser conformidade a normas da rede (Brady et al., 2021) ou expressão de identidade de grupo (Rathje et al., 2021); ver a seção 9.8.

Novidade e caça-clique. O efeito das palavras negativas pode ser específico de um veículo de caça-clique e de uma fase em que o recurso era novo para os leitores (Robertson et al., 2023).

Efeitos pequenos. Os efeitos são de 2 a 3% (Brady et al., 2021) e de cerca de 2,3% por palavra (Robertson et al., 2023); podem escalar, mas isso é argumento, e não resultado.

Alcance não é viés nem intenção. A amplificação por partido ou por fonte é dado de alcance, depende da base de rótulos e não tem mecanismo identificado (Huszár et al., 2022).

Dados que só a plataforma tem. O grupo de controle do Twitter existia para o negócio e foi revisado por equipes jurídica e de privacidade, e não por comitê de ética; o tratamento mudou ao longo do tempo (Huszár et al., 2022). Quem de fora não tem acesso aos registros não replica nem estende o estudo.

Quatro limites valem para o capítulo inteiro. Primeiro, nenhuma fonte é brasileira, e as de plataforma tratam do Twitter e do Facebook, nos Estados Unidos e, no caso de Huszár et al. (2022), em outros seis países entre os quais não está o Brasil; as plataformas mudam (o tratamento do experimento do Twitter mudou ao longo do tempo), e nada aqui diz de WhatsApp, YouTube ou TikTok. Se-

gundo, o dado de maior alcance sobre a regra de entrega (Huszár et al., 2022) mede alcance e não comportamento do produtor. Terceiro, os dados sobre conteúdo moralizado são, em maior parte, de associação, e os efeitos são pequenos. Quarto, o capítulo não mostra que o feed seja um esquema de reforçamento: mostra que, para o comportamento verbal, a intermitência é a regra e que a regra de entrega da plataforma altera o alcance; o resto é hipótese que pede dados que só as plataformas têm.

Para quem quer usar o capítulo na análise de falas (Capítulos 3, 5 e 7), a consequência é prática. O alcance de uma fala é variável que o analista do comportamento não controla e que altera o que ela pode receber; comparar a fala de um político com a de outro sem registrar o alcance de cada uma, e o que a plataforma fez com ele, mistura duas variáveis. Quando o alcance não é conhecido, o analista deve dizê-lo.

Síntese

A atenção do ouvinte é, na análise de Skinner (1957), reforçador generalizado que fortalece o comportamento verbal sem respeito à forma, e nas plataformas o que chega ao produtor é uma contagem de reações que a aproxima. A hipótese de que essa contagem seja o reforçador dominante de quem publica tem apoio parcial nos estudos do Capítulo 9 e uma cautela no de Gunnarsson et al. (2026), em que a demanda por usar redes para obter atenção foi baixa (seção 8.1). Os dados sobre conteúdo emocional e moralizado não formam um só achado: palavras negativas aumentam cliques, referência ao exogrupo aumenta compartilhamentos, e a indignação acompanha o retorno recebido; são respostas distintas, medidas por propriedades de forma, e a função que as mantém não foi medida (8.2).

A plataforma é o terceiro termo. O ranqueamento fixa o alcance de cada emissão, e a auditoria do Twitter mostra que ele difere do ordenamento cronológico, com amplificação maior para o maior partido de direita ou centro-direita em seis de sete países, nenhuma associação por político e resultados para notícias que dependem da base de rótulos; o mecanismo é desconhecido (8.3). A hipótese de que a entrega siga um esquema de razão variável é coerente com Skinner, que dá a intermitência por regra do comportamento verbal, mas nenhuma fonte a mede; a literatura dos esquemas (Ferster e Skinner, 1957) dá, para pombos, um padrão por esquema, com variedade na razão variável de médias grandes, e a comparação com o feed pede dados que só as plataformas têm. A analogia tem ainda pontos frouxos: a reação é contagem, há dois operantes candidatos e, se a forma da publicação determina o que ela recebe, o

que há é seleção de formas (8.4). Os dois experimentos com o algoritmo mudaram alcance e exposição e não mudaram atitudes declaradas; nenhum mediou o que o produtor fez depois (8.5).

Os Capítulos 3, 5 e 7 usam deste capítulo a ideia de que o alcance é variável a registrar na análise de qualquer fala; o Capítulo 3, ainda, a distinção entre a função e a forma que as propriedades do conteúdo deixam em aberto. O Capítulo 6 recebe uma resposta parcial à pergunta de quanto a regra de entrega torna permeável a fronteira da bolha: ela a move, e o que atravessa a fronteira como correção não foi medido. O Capítulo 11 (seção 11.2), ao tratar das medidas de difusão, retoma a distinção entre alcance, exposição e reação. O Capítulo 12 (seção 12.6) parte do que mudou e do que não mudou com a regra de entrega para discutir intervenções sobre as contingências da plataforma.

Referências

- Brady, W. J., McLoughlin, K., Doan, T. N., & Crockett, M. J. (2021). How social learning amplifies moral outrage expression in online social networks. *Science Advances*, 7(33), Article eabe5641. <https://doi.org/10.1126/sciadv.abe5641>
- Ceylan, G., Anderson, I. A., & Wood, W. (2023). Sharing of misinformation is habitual, not just lazy or biased. *Proceedings of the National Academy of Sciences*, 120(4), Article e2216614120. <https://doi.org/10.1073/pnas.2216614120>
- Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement*. Appleton-Century-Crofts. <https://doi.org/10.1037/10627-000>
- Guerin, B. (1994). Attitudes and beliefs as verbal behavior. *The Behavior Analyst*, 17(1), 155–163. <https://doi.org/10.1007/bf03392661>
- Gunnarsson, K. F., Rzeszutek, M. J., Malkin, A., Tewogbola, P. O., Han, S., & Jacobs, E. A. (2026). A function-based examination of social media demand. *Behavior Analysis in Practice*. Advance online publication. <https://doi.org/10.1007/s40617-026-01197-7>
- Huszár, F., Ktena, S. I., O'Brien, C., Belli, L., Schlaikjer, A., & Hardt, M. (2022). Algorithmic amplification of politics on Twitter. *Proceedings of the National Academy of Sciences*, 119(1), Article e2025334119. <https://doi.org/10.1073/pnas.2025334119>
- Lindström, B., Bellander, M., Schultner, D. T., Chang, A., Tobler, P. N., & Amodio, D. M. (2021). A computational reward learning account of social media engagement. *Nature Communications*, 12(1), Article 1311. <https://doi.org/10.1038/s41467-020-19607-x>
- Mosleh, M., & Rand, D. G. (2022). Measuring exposure to misinformation from political elites on Twitter. *Nature Communications*, 13(1), Article 7144. <https://doi.org/10.1038/s41467-022-34769-6>
- Nyhan, B., Settle, J., Thorson, E., . . . Tucker, J. A. (2023). Like-minded sources on Facebook are prevalent but not polarizing. *Nature*, 620(7972), 137–144. <https://doi.org/10.1038/s41586-023-06297-w>
- Rathje, S., Van Bavel, J. J., & van der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, 118(26), Article e2024292118. <https://doi.org/10.1073/pnas.2024292118>

Robertson, C. E., Pröllochs, N., Schwarzenegger, K., Pärnamets, P., Van Bavel, J. J., & Feuerriegel, S. (2023). Negativity drives online news consumption. *Nature Human Behaviour*, 7(5), 812–822. <https://doi.org/10.1038/s41562-023-01538-4>

Skinner, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts.



MATERIAIS EDUCACIONAIS

Por onde começar a ler Análise do Comportamento

Artigos científicos, periódicos e instituições de referência reunidos para a formação em Análise do Comportamento e em prática baseada em evidências.

Acesse, de graça e sem cadastro

walden4.com.br/materiais



9

O ciclo produtor-audiência

Hildum e Brown publicaram em 1956 um experimento em que um entrevistador telefonou para 40 estudantes homens da escola de verão de Harvard e leu 15 afirmações sobre a filosofia de Educação Geral da universidade, cada uma com quatro alternativas, de “concordo plenamente” a “discordo plenamente”. Metade deles ouviu “Good” (“bom”), em tom neutro, depois de algumas respostas; a outra metade ouviu “Mm-hmm” nos mesmos pontos. Em cada metade, dez estudantes ouviam a reação depois das respostas favoráveis à Educação Geral e dez, depois das desfavoráveis (Hildum & Brown, 1956). Com o “Good”, a média de atitude (menor é mais favorável) foi 24,95 no grupo reagido nas respostas favoráveis e 31,75 no reagido nas desfavoráveis. Dos 20 estudantes que ouviram “Good”, oito notaram que o entrevistador dizia a palavra, e só um supôs que fosse aprovação; todos negaram que as respostas tivessem sido influenciadas por ele.

A pergunta dos autores era de método: se um entrevistador pode influenciar o entrevistado sem querer, a influência opera fora da consciência de todos os envolvidos e o levantamento de opinião fica enviesado. Este capítulo faz a pergunta do lado de quem fala. O que a reação de quem ouve faz com o que se diz, e o que o que se diz faz com quem ouve? Em 1956 a reação era uma palavra de um entrevistador. Hoje é uma contagem pública de curtidas e compartilhamentos, distribuída por um algoritmo. O capítulo reúne primeiro a evidência experimental mais antiga (seção 9.1) e depois a de plataformas (9.2, 9.3, 9.5 e 9.6), propõe um esquema que as une (9.4) e uma hipótese de nível cultural (9.7), e diz em cada passo o que é dado, o que é interpretação e o que é hipótese.

9.1 Opiniões podem ser reforçadas: os experimentos de 1956 a 1969

O experimento de Hildum e Brown, que abre o capítulo, tem antecessores e sucessores. Os estudos de reforço verbal anteriores a ele tomavam como unidade de resposta uma palavra ou uma classe de palavras. Num deles, o experimentador murmurava “Mm-hmm” quando o estudante dizia um substantivo no plural; em outros, dizia “Good” depois de frases que começavam com o pronome escolhido; em todos, os sujeitos diziam não ter percebido relação entre o que diziam e a reação (Hildum & Brown, 1956). Os autores argumentam que um entrevistador de opinião reagiria a categorias de conteúdo, e não a palavras específicas, e usaram o telefone para que a reação vocal fosse a única possível.

Scott (1957) criou a situação de aprovação em sala de aula. Estudantes de 29 turmas de psicologia geral escreveram o que pensavam sobre três questões polêmicas da universidade (serviço militar universal, horário noturno das universitárias e ênfase no futebol), e cada turma recebeu uma delas para debate. Em cada turma, o instrutor escolheu duas pessoas claramente favoráveis e duas claramente contrárias e as convidou para um “experimento” sobre quanto conseguiriam influenciar a turma, com a condição de que cada uma defendesse o lado oposto ao próprio. Duas semanas depois, os pares debateram diante da turma, com três minutos de exposição e dois de réplica para cada lado, e a turma votou em cédulas qual debatedor se saíra melhor. A votação era encenada: o instrutor fingia contar as cédulas, e quem “venceria” em cada par fora definido de antemão, alternando-se o lado vencedor. Logo depois, todos reescreveram a opinião sobre a questão, e os textos, copiados em cartões identificados só por número, foram codificados numa escala de sete pontos, de muito favorável a muito contrário. Dos 58 pares designados, 36 produziram dados completos, em 19 turmas e com seis instrutores; outros 36 estudantes, com opinião clara mas sem debate, formaram o controle. Os vencedores mudaram em média 1,25 ponto na direção do lado que haviam defendido, os perdedores mudaram 0,17 ponto em sentido contrário, e os controles, 0,31 ponto no mesmo sentido dos vencedores. A diferença entre vencedores e perdedores foi estatisticamente significativa ($t = 3,97$; $p < 0,001$); o t é a estatística do teste que compara duas médias, e o p é a probabilidade de se obter uma diferença tão grande, ou maior, se na verdade os grupos não diferissem, de modo que, por convenção, p abaixo de 0,05 é chamado de “significativo”. A diferença entre vencedores e controles também foi significativa ($p < 0,02$), e a entre perdedores e controles não (Scott, 1957).

Spielberger et al. (1962) estudaram o reforço verbal numa tarefa que não é de opinião, a construção de frases, e perguntaram se o efeito exige que a pessoa perceba a contingência. Cada uma de 45 universitárias, testada individualmente, recebia 100 cartões, com um verbo no passado e os pronomes *I*, *we*, *you*, *he*, *she* e *they* em ordens diferentes, e devia compor uma frase que começasse por um dos pronomes e contivesse o verbo; cada frase era uma tentativa. Nas 20 primeiras ninguém foi reforçado, o que mede a frequência de base de frases iniciadas por *I* ou *we*. Depois, as participantes foram sorteadas para três grupos de 15, que receberam instruções neutras, instruções que pediam a primeira frase que viesse à mente, com pressa, e desestimulavam a busca de um princípio, ou instruções que diziam existir um modo correto de montar as frases, para que a percepção da contingência variasse. Nas 80 tentativas seguintes, em cada grupo 10 ouviram “good”, dito em tom neutro, depois de cada frase iniciada por *I* ou *we*, e cinco, que formavam o controle, não ouviram nada. Ao fim, uma entrevista de perguntas cada vez mais específicas, que evitava sugerir a contingência correta, classificou cada participante como consciente ou não de que o “good” seguia aquelas frases. Das 30 reforçadas, 19 foram classificadas como conscientes; só elas aumentaram o uso de *I* e *we*, e nenhuma das 11 inconscientes aumentou em mais de uma resposta (Spielberger et al., 1962).

Insko (1965) levou o procedimento para o telefone e mediu se o efeito durava. Seis assistentes de pós-graduação telefonaram, cada um, para 12 estudantes sorteados do curso de psicologia introdutória da Universidade do Havaí (72 ao todo) e leram 14 afirmações de opinião sobre a criação de uma Semana Aloha na primavera, como a de que ela traria mais dinheiro de turistas ou mais congestionamento; as afirmações não diziam que a ideia era boa ou má. O estudante respondia se concordava ou discordava de cada uma, e cada assistente disse “good” depois das respostas de atitude positiva para seis dos estudantes e depois das de atitude negativa para os outros seis. Cada resposta valia de 0 a 3 pontos, e a soma média foi 28,22 entre os reforçados nas respostas positivas e 20,69 entre os reforçados nas negativas. Cerca de uma semana depois, o experimentador, que era o professor da turma, distribuiu aos 225 alunos um questionário sobre assuntos locais, com o item sobre a Semana Aloha a dois terços do caminho. Dos 72 contactados por telefone, 70 responderam. O grupo antes reforçado nas respostas positivas ficou, em média, em 10,40, na região da escala rotulada “um pouco a favor”, e o antes reforçado nas negativas, em 8,20, na região “incerto” ($t = 2,03$; $p < 0,05$) (Insko, 1965).

Insko e Melson (1969) usaram o mesmo procedimento para perguntar se o efeito depende de a pessoa saber que está num experimento, a hipótese das características de demanda: o participante agiria como supõe que o experi-

mentador espera. No experimento I, 72 estudantes homens do curso de psicologia introdutória da Universidade da Carolina do Norte foram contatados, metade pelo procedimento habitual de recrutamento para experimentos de laboratório e metade por telefone. Em ambos os casos, o entrevistador se apresentava como membro de um centro de pesquisa por levantamentos que sondava a atitude dos estudantes sobre TV paga, e quem recebeu a ligação não foi informado de que participava de um experimento. O participante concordava ou discordava de 12 afirmações de opinião sobre TV paga, e o “good” seguia concordar com afirmações a favor e discordar de afirmações contra (condição pró) ou o inverso (condição contra); o momento do semestre, início ou fim, também foi variado. Logo depois, todos responderam a um questionário sobre a percepção do reforço, elaborado por Dulany em 1962. O experimento II repetiu o delineamento com 20 sujeitos em cada uma das quatro combinações de contexto (laboratório ou telefone) e direção do reforço, e acrescentou uma linha de base de oito afirmações sem reforço. Nos dois, o único efeito significativo foi o do reforço ($F = 6,10$ e $5,62$; o F é a estatística da análise de variância, que compara a variação entre os grupos com a variação dentro deles), sem diferença entre laboratório e telefone (Insko & Melson, 1969).

A Tabela 9.1 reúne os cinco estudos, com o que se fez, o que se observou e o que os resultados não permitem dizer. Quatro tratam de opiniões e atitudes; o de Spielberger et al. usa uma tarefa que não é de opinião e entra por ser a principal evidência contrária sobre a percepção da contingência.

TABELA 9.1

Experimentos de reforço verbal de opiniões e atitudes, de 1956 a 1969: o que se fez, o que se observou e o que não permitem dizer

Estudo	Participantes e tarefa	Reação ao que se disse	O que se observou	Ressalva
Hildum e Brown (1956)	40 estudantes homens (escola de verão de Harvard), 10 por grupo; 15 afirmações sobre a Educação Geral, por telefone	“Good” ou “Mm-hmm” depois de respostas favoráveis ou de desfavoráveis	Com “Good”, média de atitude de 24,95 (reação às favoráveis) contra 31,75 (às desfavoráveis; $t = 3,31$; $p = 0,01$); com “Mm-hmm”, diferença não significativa e em sentido oposto; replicação com 25 estudantes, mesmo padrão	8 de 20 notaram o “Good”; todos negaram influência; dos 24 que arriscaram um palpite, todos menos um disseram que o entrevistador era a favor da Educação Geral

Estudo	Participantes e tarefa	Reação ao que se disse	O que se observou	Ressalva
Scott (1957)	72 debatedores (36 pares) de 19 turmas de psicologia, defendendo a posição oposta à própria; 36 controles	Voto da turma, definido de antemão, que declarava vencedor ou perdedor	Vencedores mudaram em média 1,25 ponto (escala de 7) na direção do debate; perdedores, 0,17 em sentido oposto; controles, 0,31 (vencedores $\times$ perdedores: $t = 3,97$; $p < 0,001$)	Sem condição de debate sem voto; pós-teste em até cinco minutos; só sete vencedores inverteram a opinião; persistência não avaliada
Spielberger et al. (1962)	45 universitárias; construção de frases com pronomes (tarefa que não é de opinião); 30 reforçadas e 15 controles	“Good” depois de frases iniciadas por ou we	Só os 19 sujeitos que verbalizaram a contingência aumentaram o responder; nenhum dos 11 sem percepção subiu mais de uma resposta	Os autores ressalvam o papel da tarefa e da população; a atitude diante do reforço modulou o efeito entre os conscientes
Insko (1965)	72 estudantes (Havaí); 14 afirmações sobre uma Semana Aloha de primavera, por telefone; questionário em sala cerca de uma semana depois (N = 70)	“Good” depois de concordar ou depois de discordar	Telefone: 28,22 (reação à atitude positiva) contra 20,69 (à negativa); questionário: 10,40 contra 8,20 ($t = 2,03$; $p < 0,05$)	O autor fala em generalidade de resposta; o experimentador era o professor da turma que respondeu ao questionário
Insko e Melson (1969)	Exp. I: 72 homens; Exp. II: 80 (20 por célula); telefone ou laboratório; 12 afirmações sobre TV paga	“Good” depois de respostas pró ou contra	Efeito do reforço significativo ($F = 6,10$ e $5,62$) e igual no laboratório e por telefone; entre os que não perceberam a contingência, diferença não significativa em cada experimento; combinados, 6 de 6 comparações na direção prevista ($p \approx 0,016$)	Efeito menos robusto que em estudo anterior com as mesmas afirmações ($F = 23,09$); só 14 de 152 relataram intenção

Nota. Os números vêm dos artigos (Hildum & Brown, 1956; Scott, 1957; Spielberger et al., 1962; Insko, 1965; Insko & Melson, 1969). Em Hildum e Brown, menor nota é atitude mais favorável; em Insko, maior. “Pró” e “contra” referem-se ao tema da entrevista.

A tabela mostra um padrão com quatro partes. Primeira: nas tarefas de opinião, a aprovação de quem ouvia deslocou o que se declarou na direção reforçada, nos quatro estudos. Segunda: o efeito apareceu fora da situação em que foi

produzido. Insko (1965) o encontrou uma semana depois, em outro meio (questionário escrito) e em outro contexto (de casa para a sala de aula), e Insko e Melson (1969) o encontraram do mesmo tamanho no laboratório e por telefone, com sujeitos que não sabiam estar num experimento. Terceira: o efeito é pequeno e frágil. Em Scott (1957), só sete dos vencedores inverteram a opinião, os outros 14 que mudaram apenas a enfraqueceram, e o pós-teste veio no máximo cinco minutos depois do voto; em Insko e Melson (1969), o efeito foi bem menos robusto que em estudo anterior com as mesmas afirmações. E o reforçador importa: o “Mm-hmm” não funcionou em Hildum e Brown, que sugerem que a diferença pode vir da maior visibilidade do “Good” (p. 110) e, em parte, de seu significado mais claramente favorável. Quarta: a percepção da contingência é disputada. Em Spielberger et al. (1962), só aprenderam os sujeitos que a verbalizaram; em Insko e Melson (1969), entre os que não relataram perceber a contingência a diferença entre as condições não foi significativa em nenhum dos dois experimentos tomados separados, mas as seis comparações combinadas foram na direção prevista, e a reanálise de um estudo anterior com as mesmas afirmações deu diferença significativa. Os autores concluem que o reforço teve efeito também sem percepção e atribuem a diferença em relação à construção de frases à tarefa: opinar é resposta supra-aprendida, isto é, muito praticada, que se produz em poucas tentativas e com pouca novidade.

CONCEITO EM USO

Reforço social generalizado e audiência (Guerin, 1994), retomada do Capítulo 3

O Capítulo 3 (seção 3.3) apresentou o reforço social generalizado e a definição de Guerin (1994) de atitude como enunciado mantido por ele. Para o ciclo, interessam duas consequências. A primeira é que a consequência vem de outras pessoas e, em parte, não depende do conteúdo, como na troca de atitudes que funciona como evento social ritual, em que o reforço está na própria troca. A segunda é que a audiência é estímulo discriminativo: audiências específicas ouvem e respondem a atitudes específicas, e por isso o que se declara tende a atender às contingências da comunidade que ocasionou a declaração. O reforçador generalizado não opera igual para todos. Spielberger et al. (1962) lembram que os estudos de condicionamento verbal supõem que “Good” e “Mm-hmm” sejam reforçadores condicionados generalizados, com as mesmas propriedades para todos os sujeitos, e citam evidência de que o “Mm-hmm” elevou o desempenho de quem o avaliava como positivo e o reduziu em quem o avaliava como negativo (Mandler e Kaplan, 1956, citados por Spielberger et al., 1962).

Interpretações rivais para o deslocamento da opinião sob aprovação

Percepção da contingência. Se o falante percebe que certas respostas são aprovadas e passa a dá-las mais, o que se vê é desempenho sob regra, e não um operante moldado sem percepção. Os resultados se dividem pela tarefa (Spielberger et al., 1962; Insko & Melson, 1969), e Hildum e Brown (1956) admitem que suas medidas de percepção podem ter sido insensíveis.

Características de demanda. Se o participante sabe que está num experimento e age como supõe que se espera, o efeito deveria ser maior no laboratório que fora dele. Insko e Melson (1969) não encontraram diferença entre laboratório e telefone sem identificação como experimento.

Resposta de entrevista, e não atitude. O deslocamento pode valer só para o que se diz naquela entrevista. Insko (1965) mediu uma semana depois e em outro meio, e o próprio autor descreve o que mostrou como generalidade de resposta.

Contato com argumentos opostos. Em Scott (1957), vencedores e perdedores tiveram o mesmo contato com os argumentos opostos, e a diferença entre eles aponta para a aprovação ou a desaprovação da turma. Mas a mudança dos vencedores pode ser efeito só do reforço ou efeito do contato, que o reforço manteve, e faltou uma condição de debatedores sem voto para separar as duas leituras.

Suposição sobre o entrevistador. Em Hildum e Brown (1956), quase todos os que arriscaram um palpite supuseram que o entrevistador era a favor da Educação Geral, o que os autores atribuem à suposição sobre quem pesquisa o tema, e não ao padrão de reforço.

■ DADO

Em entrevistas por telefone (Hildum & Brown, 1956; Insko, 1965; Insko & Melson, 1969) e num debate em sala (Scott, 1957), o que as pessoas declararam como opinião deslocou-se na direção da aprovação recebida mais do que na do grupo contrário ou de controle, em efeitos pequenos e que variam com o reforçador e com a tarefa.

■ INTERPRETAÇÃO

A opinião declarada é sensível às consequências sociais que a audiência arranja para o que se declara, como a análise de Guerin (1994) prevê. A reação do ouvinte funciona como consequência do enunciado, e a audiência, como ocasião para ele.

■ HIPÓTESE

A mesma sensibilidade vale para o que se publica quando a reação da audiência é uma curtida ou um compartilhamento. Nenhum dos cinco estudos usa redes sociais nem reações públicas e contadas; a extrapolação é feita aqui.

9.2 O que o produtor recebe: curtidas, compartilhamentos e alcance

Nas plataformas, quem fala é um produtor: publica e recebe de volta, contadas e visíveis, as reações do público, o retorno (*feedback*) dos estudos. Os estudos usados aqui são de dois tipos: observacionais, que perguntam se o ato de publicar acompanha o retorno recebido, e experimentos em ambientes simulados, em que o retorno é manipulado.

Lindström et al. (2021) trataram o ato de postar como comportamento operante livre, sem tentativas discretas: o equivalente de um animal diante de uma única alavanca que pode ser pressionada a qualquer momento, com a curtida como consequência. Usaram quatro conjuntos de dados, que somam mais de um milhão de postagens de mais de 4.000 usuários: o Instagram (estudo 1, com 2.039 usuários e, em média, 418 postagens por pessoa) e três fóruns temáticos, de moda masculina, de moda feminina e de jardinagem (estudo 2, com 2.127 usuários e, em média, 91 postagens por pessoa). Os fóruns entraram porque no Instagram há motivos econômicos para postar e a possibilidade de contas fraudulentas e de curtidas falsas, que nos fóruns são menos prováveis.

Os autores fizeram duas análises. Na primeira, compararam duas formas de descrever como a taxa de postagem varia com a taxa de curtidas. A pesquisa com animais mostrou que a taxa de respostas costuma crescer com a taxa de recompensas segundo uma função hiperbólica, que sobe depressa e depois se achata, a chamada lei quantitativa do efeito, tida como assinatura do comportamento guiado por recompensa. Nos quatro conjuntos, essa função descreveu os dados melhor que uma relação linear (R^2 médio de 0,43 no estudo 1 e de 0,37 no estudo 2; o R^2 é a fração da variação dos dados que o modelo explica). Na segunda, definiram a latência entre duas postagens sucessivas (τ_{Post}) e testaram se o histórico de curtidas a previa pela causalidade de Granger, que se estabelece quando uma variável, aqui as curtidas, melhora a previsão de outra, aqui o intervalo, além do que os valores anteriores desta já preveem. As curtidas anteriores previram o intervalo até a postagem seguinte nos quatro conjuntos. Por fim, com um modelo de aprendizagem, estimaram para cada usuário a taxa média de recompensa que ele vinha recebendo. No Instagram, o intervalo entre postagens foi cerca de 18% menor, algo como oito horas, quando essa taxa estimada era alta do que quando era baixa.

O estudo 3 foi um experimento, com 176 participantes. Durante uma sessão online de 25 minutos, eles podiam postar memes quando e quantas vezes quisessem e recebiam de 0 a 19 curtidas de supostos outros usuários. Na passagem da primeira para a segunda metade da sessão, a taxa média de curtidas

mudava, de 0 a 9 por postagem para 10 a 19 em alguns participantes e no sentido inverso em outros, com a direção contrabalançada. A latência média até a postagem foi cerca de 10,9% maior quando a taxa era baixa (0 a 9) do que quando era alta (10 a 19). Os autores ressaltam que há muitos motivos para postar além da recompensa, que o estudo trata de quando se posta e não do que se posta, e que os dados não dizem se o uso intenso é mal-adaptativo.

Ceylan et al. (2023) estudaram o compartilhamento. Em quatro estudos ($N = 2.476$), pessoas com conta no Facebook viam manchetes (nos estudos 1 e 2, no formato do Facebook) e decidiam, por um botão simulado de compartilhar, se as compartilhavam: eram 16 manchetes, oito verdadeiras e oito falsas nos estudos 1, 2 e 4. A força do hábito de compartilhar notícias foi medida por quatro itens de autorrelato, numa escala de 1 (nunca) a 7 (sempre), adaptados de um índice de hábito, como “compartilhar notícias no Facebook é algo que faço automaticamente”; a média dos itens deu a nota de hábito de cada pessoa. No estudo 1 ($N = 200$), em média 32% das manchetes verdadeiras e 5% das falsas foram compartilhadas. Quem tinha o hábito mais forte (um desvio-padrão acima da média) compartilhou proporções semelhantes de manchetes verdadeiras (43%) e falsas (38%), e quem o tinha mais fraco (um desvio-padrão abaixo), 15% e 6%; os 15% mais habituais responderam por 37% das falsas compartilhadas. No estudo 2 ($N = 839$), os participantes foram sorteados para escolher primeiro se compartilhavam as 16 manchetes e depois julgar a acurácia de cada uma, ou para julgar a acurácia de todas antes de decidir se compartilhavam. Julgar a acurácia antes reduziu o compartilhamento de falsas, mas não aumentou a discriminação dos mais habituais mais do que a dos demais.

No estudo 4 (601 participantes, dos quais 120 foram excluídos por falharem na checagem de que não havia recompensa no teste, como previsto no pré-registro), cada pessoa foi sorteada para uma de três condições e fez 80 tentativas de treino: em cada uma, apertava uma tecla para compartilhar ou não uma manchete e via “ganhou +5 pontos” ou “ganhou 0 pontos”. No treino de acurácia, ganhava 5 pontos ao compartilhar uma manchete verdadeira ou ao não compartilhar uma falsa; no treino de desinformação, o inverso; no controle, não havia retorno. Em seguida decidia sobre 16 manchetes novas, oito verdadeiras e oito falsas, como faria no Facebook e sem recompensa. No teste, sem retorno, o padrão do treino persistiu: depois do treino de acurácia, 66% das verdadeiras e 24% das falsas foram compartilhadas, contra 40% e 24% no controle e 54% e 33% no treino de desinformação.

Hábito, no uso dos autores, é a resposta recompensada que se repete num contexto, associa-se a pistas recorrentes e passa a ocorrer com pouca sensibilidade aos resultados (Ceylan et al., 2023). Em termos operantes, é uma resposta

mantida por reforçamento e sob controle de estímulos recorrentes do contexto, aqui o formato da publicação e o botão de compartilhar, que persiste quando o reforçador sai de cena. O que o estudo 4 acrescenta é que a veracidade pode ser arranjada como contingência: com pontos para a acurácia, a diferença entre verdadeiras e falsas compartilhadas aumentou e persistiu sem retorno, e o treino de desinformação elevou a frequência geral de compartilhar sem derrubar a discriminação de base. Os autores concluem que o hábito emergiu como a influência principal no compartilhamento, acima do viés político e da reflexão crítica.

O sumário previa aqui monetização e votos como consequências do produtor. Nenhum estudo verificado mede receita ou voto como resultado de publicar; ficam como hipótese na seção 9.7.

9.3 O que a audiência faz e o que a mantém

A audiência também se comporta, e é nela que nasce o reforço do produtor. O que os estudos verificados mostram é, sobretudo, o que ela faz diante de cada tipo de conteúdo. Rathje et al. (2021) analisaram 2.730.215 publicações do Facebook e do Twitter. No estudo 1, eram contas de veículos de notícia classificados como liberais (por exemplo, New York Times e MSNBC) ou conservadores (por exemplo, Fox News e Breitbart), com 599.999 posts do Facebook e 227.229 tweets; no estudo 2, que replicou o primeiro em outro contexto, eram contas de membros do Congresso dos Estados Unidos, democratas e republicanos, com 1.078.562 tweets (até 3.200 de cada parlamentar) e 825.424 posts do Facebook. Os dados foram coletados em 2020.

Para saber se uma publicação falava do próprio campo político (o endogrupo) ou do oposto (o exogrupo), os autores contaram, em cada texto, as palavras de três listas: os 100 políticos democratas e republicanos mais famosos segundo o YouGov, todos os parlamentares dos dois partidos e cerca de dez termos de identidade política (como “liberal” e “conservative”). Um tweet com duas palavras da lista republicana, por exemplo, recebia nota 2 nessa categoria. Numa conta conservadora, as palavras conservadoras contavam como linguagem do endogrupo e as liberais, como linguagem do exogrupo; nas contas liberais valia o inverso. Pelo mesmo método contaram palavras de afeto negativo, de afeto positivo e de linguagem moral-emocional, com dicionários já validados em outros estudos, para comparar o peso da linguagem do exogrupo com o dessas emoções. Em cada conjunto de dados, uma regressão relacionou o número de compartilhamentos (em logaritmo) a essas contagens, descontados o efeito de ter link, foto ou vídeo, o número de seguidores da conta e o fato de o tweet ser

um retuíte; as mesmas regressões foram aplicadas a cada reação do Facebook. O resultado de cada regressão é um coeficiente por palavra, que os autores expressam como aumento percentual.

Cada termo referente ao exogrupo se associou a um aumento de 35 a 57% na difusão nas contas de veículos de notícia e de 65 a 180% nas do Congresso, conforme o conjunto. A média dos oito conjuntos, numa meta-análise de efeitos aleatórios feita pelos autores, foi de cerca de 67% por palavra (41% nos veículos e 99% no Congresso), efeito cerca de 4,8 vezes o da linguagem de afeto negativo e 6,7 vezes o da linguagem moral-emocional. Os autores falam em aumento da chance de a publicação ser compartilhada; o modelo estima, mais exatamente, a razão entre os números esperados de compartilhamentos. As publicações que se referiam ao exogrupo foram compartilhadas cerca de duas vezes mais que as que se referiam ao endogrupo. O Facebook oferece seis reações (“like”, “love”, “haha”, “sad”, “wow” e “angry”): a “angry” foi a mais popular em sete dos oito conjuntos, na contagem dos autores, e a linguagem do exogrupo previu reações “angry”, a do endogrupo, reações “love”. O efeito do exogrupo não foi moderado por orientação política nem por plataforma, e foi mais forte entre líderes políticos que entre veículos de notícia.

Os autores leem o resultado em termos de identidade social: mensagens que satisfazem motivos de identidade de grupo podem receber mais engajamento, e publicações sobre o exogrupo podem ter sucesso por apelarem a raiva, indignação e zombaria (Rathje et al., 2021). É uma leitura de outro referencial. Em termos operantes, o que o estudo estabelece é que certas propriedades do conteúdo, aqui a referência ao exogrupo, se associam a mais respostas de compartilhar e de reagir; por que essas respostas são reforçadas para quem as emite é o que ele não mede.

Guerin (1994) oferece uma hipótese vizinha para essa parte do ciclo. Trocar atitudes pode funcionar como evento social ritual, em que o reforço está na própria troca e não no que se diz, e a audiência é estímulo discriminativo, de modo que o que se declara tende a atender às contingências da comunidade que ocasionou a declaração. Para a audiência de uma rede, isso sugere que curtir, compartilhar ou comentar pode ser, antes de uma avaliação do conteúdo, parte da troca que mantém a comunidade. É hipótese: os estudos citados não medem a consequência que a própria audiência recebe.

■ DADO	A taxa de curtidas se associa a menor intervalo entre postagens (Lindström et al., 2021); o padrão de compartilhar, instalado por retorno, persiste quando o retorno é retirado (Ceylan et al., 2023); publicações sobre o exogrupo político são compartilhadas cerca de duas vezes mais que as sobre o endogrupo (Rathje et al., 2021).
■ INTERPRETAÇÃO	Publicar e compartilhar são operantes sensíveis ao retorno social e ao contexto da plataforma, e o conteúdo a que a audiência mais responde é, em parte, o que o produtor passa a produzir.
■ HIPÓTESE	O compartilhar e o reagir da audiência são mantidos pelo retorno que ela recebe da própria rede. Nenhum dos estudos o mede.

9.4 O ciclo: cada lado modela o outro

O ciclo é o encaixe de duas contingências de três termos, a do produtor (seção 9.2) e a da audiência (seção 9.3). A do produtor tem como ocasião a audiência, a norma da rede e a interface da plataforma; como resposta, publicar um conteúdo; como consequência, curtidas, compartilhamentos e alcance. A da audiência tem como ocasião o conteúdo que aparece no feed; como resposta, curtir, compartilhar, comentar ou ignorar; e, como consequência, o que os estudos não medem e a seção 9.3 deixou como hipótese. O que as faz um ciclo é que cada uma contém, como termo, a resposta da outra: a publicação é a ocasião da audiência, e a reação da audiência é a consequência do produtor. É o que Glenn (2004) chama de contingências comportamentais entrelaçadas.

CONCEITO EM USO

Contingências comportamentais entrelaçadas (Glenn, 2004), retomada do Capítulo 5

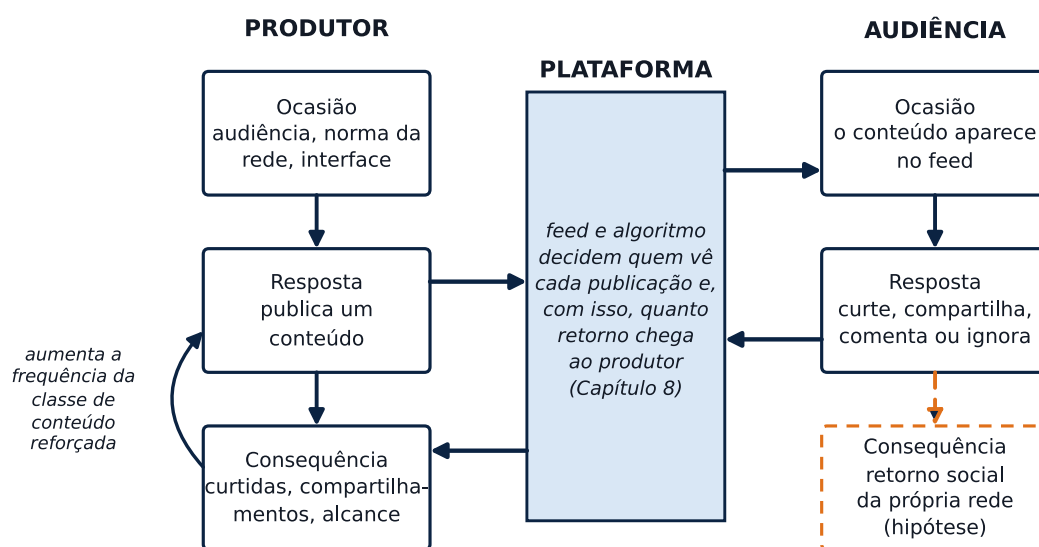
O Capítulo 5 (seção 5.1) apresentou o conceito: as contingências estão entrelaçadas quando o comportamento de duas ou mais pessoas funciona como evento ambiental para o comportamento das outras (Glenn, 2004), e o resultado não é o efeito cumulativo de cada pessoa se comportando sozinha, e sim o de seu comportamento inter-relacionado. No Capítulo 5 os participantes eram falante, audiência e grupo-alvo; aqui são o produtor e a audiência, e entre eles está a plataforma.

A figura não é um achado, é um modelo. Duas de suas setas têm dados nos estudos do capítulo: a que vai do retorno à produção (Lindström et al., 2021; Ceylan et al., 2023; Brady et al., 2021, seção 9.5) e a que liga o tipo de conteúdo à

reação da audiência (Rathje et al., 2021). A que fecha o ciclo, em que o que a audiência vê muda o que ela passa a fazer, tem apoio indireto: Brady et al. (2021) mostram, em ambiente simulado, que o que aparece no feed funciona como norma e que a norma altera a expressão (seção 9.5). Nenhum estudo mede, na mesma população e ao longo do tempo, as duas metades. A faixa do meio é a plataforma, que decide quem vê o quê e, portanto, quanto retorno chega a quem publica; as fontes verificadas dizem pouco sobre ela, e o Capítulo 8 trata dela. Aqui ela é o termo médio.

FIGURA 9.1

O ciclo produtor-audiência: duas contingências de três termos que se entrelaçam



A publicação é a ocasião da audiência; a reação da audiência é a consequência do produtor. Linha contínua: relação com dados nos estudos do capítulo. Tracejada: hipótese.

Nota. Esquema original do autor, sem dados. Linha contínua: relações com dados nos estudos do capítulo; tracejada: hipótese. A plataforma decide quanto retorno chega ao produtor (Capítulo 8).

9.5 Indignação moral aprendida por reforço: a evidência

O estudo que mais se aproxima do ciclo da seção 9.4, entre os lidos, é o de Brady et al. (2021), sobre a expressão de indignação moral. Os autores partem de duas hipóteses: que as pessoas mudam o que expressam por aprendizagem por reforço, alterando comportamentos expressivos em resposta ao retorno social, e por aprendizagem de normas, ajustando a expressão ao que inferem ser normativo entre os pares, por observação. O delineamento tem duas partes.

Na primeira, dois estudos observacionais pré-registrados reuniram as histórias completas de tweets de usuários “politicamente engajados”, identificados por terem tuitado ao menos uma vez sobre as audiências de confirmação de

um juiz da Suprema Corte, em outubro de 2018 (estudo 1, 3.669 usuários), ou sobre um incidente em que uma companhia aérea maltratou um passageiro (estudo 2, a mesma quantidade); os autores contam 7.331 usuários no total, com 12,7 milhões de tweets. Para dizer se um tweet expressava indignação moral, definiram-na por três componentes (sentimentos diante da violação percebida de princípios morais pessoais; emoções como raiva, nojo e desprezo; e reações como culpar, responsabilizar ou querer punir) e treinaram um classificador em 26.000 tweets anotados com base nessa definição. O classificador teve acurácia de 75% (proporção de tweets classificados corretamente) e *F1* de 0,71 (nota de 0 a 1 que combina quantos dos tweets de indignação ele achou e quantos dos que marcou como indignação eram mesmo).

Depois, para cada usuário, relacionaram a quantidade de indignação de um dia ao retorno (curtidas e compartilhamentos) que a indignação do dia anterior havia recebido, descontados a frequência de tuitar, o número de seguidores, a presença de links e de mídia, o histórico da semana anterior e o retorno que não era específico à indignação (para separar o efeito do retorno à indignação do simples tuitar mais quando se recebe mais retorno). A expressão diária de indignação associou-se positivamente ao retorno recebido pela indignação do dia anterior (coeficientes de regressão *b* de 0,03 e 0,02 nos estudos 1 e 2). No exemplo dos autores, quem tinha em média cinco curtidas ou compartilhamentos por tweet e recebeu dez ao expressar indignação (um aumento de 100%) tenderia a aumentar a indignação do dia seguinte em 2 a 3%, efeito que os autores chamam de pequeno e capaz de escalar. Quanto mais longa a história do usuário, menor o efeito do retorno recente, e usuários inseridos em redes ideologicamente mais extremas expressaram mais indignação e responderam menos ao retorno.

Na segunda parte, dois experimentos pré-registrados (120 participantes cada, recrutados no Prolific, todos liberais) simularam um Twitter. No estudo 3, os participantes eram sorteados para uma de duas condições de norma e começavam rolando um feed de 12 tweets de uma “nova rede”, cujas preferências deveriam aprender. Na condição de norma de indignação, 75% dos tweets expressavam indignação e exibiam mais curtidas e compartilhamentos que os neutros; na de norma neutra, todos eram neutros. Em seguida faziam 30 tentativas em que escolhiam, entre dois tweets políticos, um de indignação e um neutro, para “postar” na rede, e recebiam curtidas que, segundo lhes diziam, vinham de participantes anteriores; o objetivo era maximizar essas curtidas. Escolher indignação rendia mais curtidas, em média, nas duas condições. O estudo 4 usou o mesmo paradigma com uma diferença: o retorno maior ia para o tweet congruente com a norma, isto é, para a indignação na condição de norma de indignação e para o tweet neutro na de norma neutra.

Os autores mediram dois processos. A aprendizagem de norma é a tendência de escolher, na primeira tentativa, o tipo de tweet que o feed mostrava ser o normal. A aprendizagem por reforço é a tendência de escolher cada vez mais, ao longo das tentativas, o tipo de tweet que rendia mais curtidas. Os resultados vêm como razões de chances (OR, de *odds ratio*): OR de 1 significa chances iguais, e acima de 1, vantagem para a escolha em questão. Na primeira tentativa, quem viu a norma de indignação escolheu o tweet de indignação com chance muito maior (OR = 4,94 no estudo 3 e 5,38 no estudo 4), e quem viu a norma neutra escolheu o neutro (OR = 1,73 e 1,54). Ao longo das tentativas, a escolha do tweet reforçado aumentou (no estudo 3, OR = 1,10 no conjunto das duas condições; no estudo 4, 1,03 para indignação e 1,06 para tweets neutros). E o efeito do reforço foi menor entre quem viu a norma de indignação (OR de 0,95 no estudo 3 e de 0,97 no estudo 4 para a diferença entre as condições; abaixo de 1, o efeito é menor nessa condição), o que os autores leem como menor dependência do retorno para quem já via a indignação como norma.

Os termos dos autores vêm da aprendizagem por reforço computacional (retorno, erro de predição). Importa aqui a relação funcional: a expressão de indignação variou com as consequências sociais que a antecederam, em dados naturais e num ambiente em que o retorno foi arranjado. O que o reforço sozinho não explica, dizem os autores, é que o feed mostra o que é normativo na rede e quem o observa depende menos do retorno. Em linguagem analítico-comportamental, isso se aproxima de controle por modelo e por regra (leitura aqui proposta), e não é o mesmo que reforçamento direto. A ligação com os algoritmos é interpretação dos autores: por determinarem quantos usuários veem uma publicação, os algoritmos afetam o retorno e podem influenciar comportamentos morais ao explorar tendências de aprendizagem por reforço; e decisões de desenho voltadas a lucro por engajamento podem afetá-lo indiretamente, porque o conteúdo que provoca indignação atrai engajamento alto. Nenhum dos quatro estudos mede decisões de desenho.

Os próprios autores registram os limites: os estudos observacionais não permitem inferência causal, os usuários foram identificados por terem tuitado em episódios de indignação pública, e as conclusões dizem respeito à expressão em texto, e não à emoção (seção 9.8).

■ DADO	Em 12,7 milhões de tweets de 7.331 usuários, a indignação do dia seguinte aumentou de 2 a 3% para cada 100% de aumento do retorno à indignação do dia anterior; em dois experimentos de 120 participantes, a escolha de tweets de indignação aumentou com o retorno e acompanhou a norma exibida (Brady et al., 2021). Em mais de um milhão de postagens, o intervalo entre postagens diminuiu com a taxa de curtidas (Lindström et al., 2021).
■ INTERPRETAÇÃO	A expressão de indignação, como a frequência de postar, é sensível ao retorno social; o contexto da rede, observado no feed, modula essa sensibilidade.
■ HIPÓTESE	Em plataformas em que o algoritmo amplia o que provoca engajamento, o retorno por indignação cresce e a expressão cresce com ele. É hipótese dos autores, retomada aqui; nenhum estudo manipulou o algoritmo e mediu a indignação.

9.6 O político como produtor de conteúdo

Um político que publica em rede é um produtor com uma audiência particular: seguidores com história própria de reforço sobre o tema, eleitores potenciais, imprensa. Rathje et al. (2021) separam a análise por tipo de conta e por campo político, o que permite ler o ciclo para esse produtor. Em 1.078.562 tweets e 825.424 publicações no Facebook de congressistas democratas e republicanos, as que se referiam ao exogrupo foram mais compartilhadas nos dois campos. O aumento estimado por termo sobre o exogrupo, na meta-análise dos autores, foi de 99% nos conjuntos de congressistas, contra 41% nos de veículos de notícia, e o efeito não foi moderado por orientação política. O estudo cobre os anos que antecederam a eleição de 2020, é correlacional, e os autores dizem não saber como as escolhas algorítmicas contribuem.

O achado é sobre contas e conteúdo: diz o que recebeu mais compartilhamento, e não por que cada político publicou o que publicou. O estudo aplica o mesmo critério aos dois campos, que é o que o método seguido aqui pede. Aplicá-lo a políticos brasileiros é a pergunta que se quer fazer; nenhuma das fontes verificadas a responde.

■ DADO	Publicações de congressistas dos Estados Unidos que se referiram ao exogrupo político foram mais compartilhadas que as que se referiam ao endogrupo, no Facebook e no Twitter, nos dois campos, com efeito maior que o observado em veículos de notícia (Rathje et al., 2021).
■ INTERPRETAÇÃO	O retorno que a audiência dá a um produtor político favorece, nos dados, o conteúdo sobre o exogrupo, o que descreve contingências do produtor e não julga pessoas ou partidos.
■ HIPÓTESE	O mesmo vale para políticos brasileiros e para outras plataformas. Nenhuma fonte verificada o testa.

9.7 A metacontingência do ecossistema: produtor, audiência, plataforma e anunciante

Até aqui o capítulo ficou no nível do comportamento de indivíduos em contingências entrelaçadas. A pergunta que resta é se o conjunto, formado por produtores, audiências, plataformas e quem paga por atenção, se mantém por si, como uma prática que recorre porque seus efeitos a selecionam. A Análise do Comportamento tem um conceito para isso, a metacontingência. O capítulo o usa como hipótese e com os cuidados que a própria literatura pede.

CONCEITO EM USO

Metacontingência (Glenn, 1988, 2004)

Metacontingência é a unidade de análise que engloba uma prática cultural, em todas as suas variações, e o resultado agregado das variações correntes (Glenn, 1988). Os resultados culturais não selecionam o comportamento de indivíduos; selecionam as contingências comportamentais entrelaçadas que compõem a prática. Em Glenn (2004), as linhagens que evoluem são essas contingências recorrentes, que funcionam como unidade integrada e têm um resultado que afeta a probabilidade de sua recorrência futura. A fórmula “contingências entrelaçadas, seu produto agregado e seu sistema receptor” é de Glenn e Malott (2004, citados por Houmanfar & Rodrigues, 2006). Macrocontingência é outra coisa: a relação entre uma prática e a soma agregada das consequências do macrocomportamento que a constitui, efeitos cumulativos que não podem controlar, como consequências operantes, o comportamento de indivíduos (Glenn, 2004) e que não envolvem seleção em nível cultural.

Aplicado ao ecossistema, o esquema exige três coisas: contingências entrelaçadas recorrentes, um produto agregado e um sistema receptor que, ao receber o produto, mantém a recorrência do que o gerou. A Figura 9.2 propõe o

preenchimento. As contingências entrelaçadas são as da seção 9.4, com a plataforma como terceiro termo: o produtor publica e recebe reação, a audiência vê e reage, a plataforma ordena o feed e retém atenção. O produto agregado seria um padrão recorrente de conteúdo e de engajamento, por exemplo publicações sobre o exogrupo com muito compartilhamento. O sistema receptor seria quem paga pela atenção retida: anunciantes e, por meio deles, a receita da plataforma.

FIGURA 9.2

Hipótese: a metacontingência do ecossistema produtor-audiência-plataforma



Termos: contingências entrelaçadas → produto agregado → sistema receptor (Glenn e Malott, 2004, citados por Houmanfar e Rodrigues, 2006).

Nota. Esquema original do autor, sem dados: é uma hipótese, não um resultado. Os três termos seguem a fórmula de Glenn e Malott (2004, citados por Houmanfar & Rodrigues, 2006). Houmanfar e Rodrigues propõem outra configuração: ambiente cultural, produto agregado, sistema receptor.

As fontes verificadas sustentam as contingências entrelaçadas em parte (seções 9.2 a 9.5) e deixam a segunda metade da hipótese como pressuposto de seus autores: Brady et al. (2021) afirmam que o conteúdo que provoca indignação atrai engajamento alto, o que pode afetar o comportamento moral quando o desenho visa ao lucro por engajamento; Rathje et al. (2021), que o modelo de negócios das redes se baseia em gerar engajamento, o que leva à receita de publicidade; Ceylan et al. (2023), que usuários habituais são integrais aos modelos de lucro baseados em anúncios. São premissas de interpretação, e não resultados: nenhum desses estudos mediu receita nem decisão de desenho.

Dois cuidados vêm da própria literatura. O primeiro é a distinção entre metacontingência e macrocontingência. Se o que se quer explicar é um efeito cumulativo, como a polarização (exemplo aqui escolhido), o que há é uma macrocontingência: o efeito agregado de muitos operantes independentes não

pode funcionar como consequência que controla o comportamento de cada indivíduo (Glenn, 2004). O que mantém o usuário postando continua sendo o retorno que ele mesmo recebe. A engenharia cultural, escreve Glenn (2004), exige identificar as macrocontingências que produzem efeitos indesejáveis e alterar as contingências operantes ou as metacontingências relevantes; a Parte IV volta a isso. O segundo cuidado é que o conceito é discutido. Houmanfar e Rodrigues (2006) propõem a configuração ambiente cultural, produto agregado e sistema receptor, com o produto agregado como o que é selecionado, porque julgam que a forma atual mistura níveis de análise, e dizem que as ideias são preliminares. Para o ecossistema a diferença importa: se o que o sistema receptor seleciona é o produto agregado, e não as contingências entrelaçadas, a pergunta empírica é quais produtos recorrem (que tipo de conteúdo e de engajamento), e não quais arranjos de produtor, audiência e plataforma.

■ DADO	Os estudos verificados mostram sensibilidade ao retorno social ao publicar, compartilhar e expressar indignação (seções 9.2 a 9.5) e mais compartilhamento de conteúdo sobre o exogrupo (9.3 e 9.6). Seus autores afirmam, sem medir, que os modelos de negócio das plataformas dependem de engajamento.
■ INTERPRETAÇÃO	As contingências do produtor e da audiência são entrelaçadas no sentido de Glenn (2004), com a plataforma como mediadora do retorno.
■ HIPÓTESE	O padrão entrelaçado recorre porque seu produto agregado é recebido por anunciantes que pagam por atenção. Isso exigiria mostrar que a variação do produto agregado é seguida de variação na receita e que esta mantém a prática. Nenhuma fonte verificada o faz.

9.8 Interpretações rivais e limites

O ciclo tem leituras rivais, e o que separa uma da outra depende do tipo de dado. Os dados de plataforma são de dois tipos, com limites diferentes. Os observacionais (Lindström et al., 2021; Brady et al., 2021, estudos 1 e 2; Rathje et al., 2021) descrevem associações em populações autosselecionadas; os experimentais (Lindström et al., 2021, estudo 3; Brady et al., 2021, estudos 3 e 4; Ceylan et al., 2023) manipulam o retorno em ambientes simulados, com participantes de plataformas de pesquisa, e medem escolhas, não publicações em redes reais.

Quatro limites valem para o capítulo inteiro. Primeiro, nenhuma fonte é brasileira, e as de plataforma são dos Estados Unidos ou de amostras de pesquisa online; Brady et al. (2021) dizem não saber se os achados generalizam para outras populações, plataformas ou fora do contexto político dos Estados Unidos, e que usuários do Twitter não são representativos. Segundo, os efeitos medidos são pequenos: 2 a 3% de aumento na indignação do dia seguinte para cada 100% de aumento do retorno (Brady et al., 2021) e diferença de cerca de 10,9% na latência (Lindström et al., 2021). Efeitos pequenos podem escalar, como notam os autores, mas isso é argumento, e não resultado. Terceiro, os experimentos de 1956 a 1969 usam estudantes e, em Scott (1957), um intervalo de cinco minutos. Quarto, o ciclo como um todo é um modelo: cada seta tem apoio parcial, e o fechamento do ciclo ao longo do tempo, com produtores e audiências reais, não foi medido.

Interpretações rivais para o mesmo conjunto de dados

Exposição determinada pelo algoritmo. O que uma publicação recebe depende de quantos usuários a veem, e isso é decidido pela plataforma (Brady et al., 2021). Uma associação entre conteúdo e reações pode refletir o que o algoritmo mostrou, e não o que a audiência escolheria; Rathje et al. (2021) dizem não saber como as escolhas algorítmicas contribuem.

Seleção do produtor e da rede. Usuários que expressam mais indignação podem seguir mais contas extremas, em vez de aprender a expressá-la na rede em que estão; os autores admitem que a composição da rede pode ser endógena (Brady et al., 2021, estudos 1 e 2).

Outros motivos para postar. Há muitos motivos além da recompensa, como auto-expressão e desenvolvimento de relações (Lindström et al., 2021).

Norma e identidade em lugar de reforço. O que se observa pode ser conformidade a normas da rede (Brady et al., 2021) ou expressão de identidade de grupo (Rathje et al., 2021), sem que o retorno individual seja o que mantém a resposta.

Expressão sem emoção. As conclusões dizem respeito à expressão em texto, e não à emoção, que não foi medida; os autores admitem que o desenho das plataformas pode desacoplar uma da outra (Brady et al., 2021).

Do telefone à rede. Os experimentos de 1956 a 1969 usam um ouvinte conhecido, uma reação verbal e poucas interações; as redes têm audiências múltiplas, retorno contado e visível e ciclos longos. A passagem é por analogia, e aqui ela é assumida como tal.

Para quem quer usar o capítulo na análise de falas (Capítulos 3, 5 e 7), a consequência é prática. A reação de quem ouve é variável de controle plausível e mensurável do que o falante diz, mas afirmar que a fala de um político específico foi mantida por determinada audiência exige o registro dessa audiência e de suas reações, que uma checagem, por si, não traz.

Síntese

O que se diz é sensível ao que acontece depois de ser dito, e isso foi mostrado com opiniões declaradas antes de existirem redes sociais: com “Good” depois de respostas, com um voto de vitória depois de um debate. O efeito é pequeno, aparece fora do laboratório, depende do reforçador e da tarefa, e a percepção da contingência é disputada (seção 9.1). Nas plataformas, a frequência de postar acompanha a taxa de curtidas, o compartilhamento se torna pouco sensível à veracidade quando se instala por repetição e a expressão de indignação acompanha o retorno e a norma exibida no feed, com efeitos pequenos nos dados observacionais (seções 9.2 e 9.5). Do lado da audiência, sabe-se o que ela faz diante de cada conteúdo; a consequência que a mantém é hipótese (9.3). O ciclo une as duas metades como contingências entrelaçadas, tem apoio parcial em cada seta e nenhum estudo o mede por inteiro (9.4).

Para o político como produtor, o único dado verificado é o de congressistas dos Estados Unidos: o conteúdo sobre o exogrupo é mais compartilhado, nos dois campos, e mais entre congressistas que entre veículos (9.6). A passagem ao Brasil é hipótese. Em nível cultural, o ecossistema pode ser lido como metacontingência, com plataforma e anunciantes como termos, mas isso exige um resultado que nenhuma fonte verificada mede, e parte do que se atribui a ele, como a polarização, é efeito cumulativo, de macrocontingência (9.7).

Os Capítulos 3, 5 e 7 usam deste capítulo a ideia de que a reação da audiência é variável de controle da fala; o Capítulo 8, a faixa da plataforma; o Capítulo 12, a distinção entre agir sobre as contingências operantes do produtor e da audiência e agir sobre as metacontingências.

Referências

- Brady, W. J., McLoughlin, K., Doan, T. N., & Crockett, M. J. (2021). How social learning amplifies moral outrage expression in online social networks. *Science Advances*, 7(33), Article eabe5641. <https://doi.org/10.1126/sciadv.abe5641>
- Ceylan, G., Anderson, I. A., & Wood, W. (2023). Sharing of misinformation is habitual, not just lazy or biased. *Proceedings of the National Academy of Sciences*, 120(4), Article e2216614120. <https://doi.org/10.1073/pnas.2216614120>

- Glenn, S. S. (1988). Contingencies and metacontingencies: Toward a synthesis of behavior analysis and cultural materialism. *The Behavior Analyst*, 11(2), 161–179. <https://doi.org/10.1007/bf03392470>
- Glenn, S. S. (2004). Individual behavior, culture, and social change. *The Behavior Analyst*, 27(2), 133–151. <https://doi.org/10.1007/bf03393175>
- Guerin, B. (1994). Attitudes and beliefs as verbal behavior. *The Behavior Analyst*, 17(1), 155–163. <https://doi.org/10.1007/bf03392661>
- Hildum, D. C., & Brown, R. W. (1956). Verbal reinforcement and interviewer bias. *The Journal of Abnormal and Social Psychology*, 53(1), 108–111. <https://doi.org/10.1037/h0040518>
- Houmanfar, R., & Rodrigues, N. J. (2006). The metacontingency and the behavioral contingency: Points of contact and departure. *Behavior and Social Issues*, 15(1), 13–30. <https://doi.org/10.5210/bsi.v15i1.342>
- Insko, C. A. (1965). Verbal reinforcement of attitude. *Journal of Personality and Social Psychology*, 2(4), 621–623. <https://doi.org/10.1037/h0022485>
- Insko, C. A., & Melson, W. H. (1969). Verbal reinforcement of attitude in laboratory and nonlaboratory contexts. *Journal of Personality*, 37(1), 25–40. <https://doi.org/10.1111/j.1467-6494.1969.tb01729.x>
- Lindström, B., Bellander, M., Schultner, D. T., Chang, A., Tobler, P. N., & Amodio, D. M. (2021). A computational reward learning account of social media engagement. *Nature Communications*, 12(1), Article 1311. <https://doi.org/10.1038/s41467-020-19607-x>
- Rathje, S., Van Bavel, J. J., & van der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, 118(26), Article e2024292118. <https://doi.org/10.1073/pnas.2024292118>
- Scott, W. A. (1957). Attitude change through reward of verbal behavior. *The Journal of Abnormal and Social Psychology*, 55(1), 72–75. <https://doi.org/10.1037/h0040249>
- Spielberger, C. D., Levin, S. M., & Shepard, M. G. (1962). The effects of awareness and attitude toward the reinforcement on the operant conditioning of verbal behavior. *Journal of Personality*, 30(1), 106–121. <https://doi.org/10.1111/j.1467-6494.1962.tb02308.x>



TV WALDEN4

Ciência em vídeo, no ar o dia inteiro

Aulas, palestras e conversas sobre psicologia, autismo e Análise do Comportamento, em canais temáticos que tocam sem parar, sincronizados para todo mundo.

667 vídeos em 10 canais

Acesse, de graça e sem cadastro

walden4.com.br/tv



10

A janela de Overton: o que se pode dizer sem ser punido

Um experimento de Bursztyn et al. (2020), feito em 2018, serve de entrada para a janela de Overton. Os 1.600 moradores da região metropolitana de Pittsburgh que participaram receberam a mesma oferta: um bônus em dinheiro se autorizassem uma doação, em seu nome, a uma organização fortemente anti-imigração. Ao acaso, uns souberam que Trump havia ganho na área metropolitana na eleição de 2016; outros, que Clinton havia levado o condado em que fica Pittsburgh. Metade de cada grupo foi levada a crer que a decisão seria anônima; a outra metade, que seria visível a outros moradores da região. Na condição anônima, aceitaram 31% (Trump ganhou) e 30% (Clinton ganhou). Na condição visível, 33% e 20% (Bursztyn et al., 2020).

A informação que variou era uma só e não dizia nada sobre imigração: dizia quem ganhara a eleição na vizinhança. O que variou com ela foi o que as pessoas aceitaram fazer quando outras poderiam ver. É esse deslocamento do que se pode fazer ou dizer em público sem custo que a conversa política chama de deslocamento da janela de Overton.

O capítulo pergunta o que o conceito significa em termos de contingências e o que os dados verificados permitem dizer sobre quem desloca a janela e por quê. A seção 10.1 apresenta o conceito, a origem e a tradução aqui proposta; a 10.2 o põe na relação entre falante e audiência; a 10.3 volta ao experimento de Pittsburgh; a 10.4 trata de duas eleições, a brasileira de 2018 e a americana de 2016; a 10.5, de políticos e redes sociais; a 10.6, do que mantém a borda no lugar; a 10.7 propõe perguntas para situar uma fala; e a 10.8 reúne as interpretações rivais.

Três cautelas ao ler o capítulo. A primeira é que o conceito nasceu num centro de pesquisa que o usa para promover as próprias políticas e é, na origem, uma metáfora sobre políticas públicas, e não uma definição de comportamento verbal. A segunda é que os dados medem coisas diferentes (percepção, ação pública, texto classificado), e nenhum estudo verificado liga a fala de um político à fala de seus eleitores com um desenho que identifique a direção da influência. A terceira é que os estudos de eleição tratam das vitórias de Trump, em 2016, e de Bolsonaro, em 2018; nenhum trata de um caso espelhado, a vitória de candidato do campo oposto. O critério do capítulo vale para qualquer campo, mas o que se mostrou não permite afirmar simetria. Em todas as seções, dado, interpretação e hipótese ficam separados.

10.1 O conceito, a origem e a tradução proposta

A formulação de origem é a do Mackinac Center for Public Policy, que mantém uma página sobre o conceito. Segundo ela, a janela de Overton é um modelo para entender como as ideias da sociedade mudam com o tempo e influenciam a política. O núcleo é que os políticos são limitados no que podem apoiar: em geral só adotam políticas amplamente aceitas como legítimas, e essas estão dentro da janela; as demais existem, mas quem as defende arrisca perder apoio popular, e estão fora dela (Mackinac Center, 2019).

A janela pode se deslocar e se expandir. Às vezes os políticos a movem ao endossar uma política que está fora dela, mas isso é raro; com mais frequência a janela se move pela lenta evolução dos valores e das normas da sociedade, que não se controla facilmente de cima. A conclusão da página é que os políticos são mais seguidores que líderes, e que as instituições que estabelecem e reforçam normas (famílias, locais de trabalho, amigos, mídia, igrejas, associações, escolas) importam mais para moldar a política do que se costuma reconhecer. O exemplo é a Lei Seca: política federal algumas gerações atrás, hoje nenhum político defende restabelecê-la, e ela saiu da janela. Interessa aqui uma passagem das perguntas frequentes: a janela reflete o que a sociedade acredita, e isso pode ser tão influenciado por fatos quanto por informação imprecisa ou enganosa; a página responde que a janela pode ser deslocada por mentiras, distorções ou mal-entendidos, e acrescenta que é obviamente errado disseminar intencionalmente informação enganosa, e dá como exemplos de erros que deslocaram a janela a subestimação dos custos do Medicare e a crença nas armas de destruição em massa do Iraque (Mackinac Center, 2019).

O leitor deve ler isso sabendo de onde vem. O conceito foi desenvolvido em meados da década de 1990 por Joseph P. Overton, que era vice-presidente sênior do centro quando morreu, em 2003, e o centro diz usar a janela de Over-

ton para ajudar a promover as próprias políticas. A página admite, por fim, que o modelo não descreve tudo sobre como a política funciona: descreve uma coisa, que os políticos só defendem políticas que julgam não prejudicar suas chances eleitorais (Mackinac Center, 2019). A descrição não fica invalidada por isso, mas é de uma parte interessada e não de um estudo, e o capítulo a usa como ponto de partida, e não como evidência.

Como formulação para uma ciência do comportamento, o conceito tem uma falta e uma troca a fazer. A falta é operacional: a página não diz como medir “amplamente aceita” nem “apoio popular”, de modo que, tal como formulado, o que está dentro ou fora da janela só se conhece depois, pelo que os políticos fizeram. A troca é de unidade e de critério. A unidade deixa de ser a política pública que um político pode apoiar e passa a ser o enunciado que um falante pode emitir. O critério deixa de ser o “apoio popular” e passa a ser a consequência que o enunciado recebe das audiências que importam para o falante, para a qual a análise do comportamento verbal já tem conceito: a audiência que pune. O quadro reúne o que Skinner (1957) diz.

CONCEITO EM USO

Audiência negativa e audiência não punitiva (Skinner, 1957)

Skinner (1957) distingue o ouvinte que apenas não reforça o comportamento verbal daquele que o pune: uma audiência em cuja presença o comportamento verbal é punido pode ser chamada de “audiência negativa”, e reis, altos funcionários e executivos poderosos podem ser audiências negativas efetivas. Há audiências negativas para assuntos (aprende-se a não mencionar certos tópicos ou eventos) e audiências que controlam só parte do repertório do falante, como a comunidade que pune a língua falada no lar da criança. O efeito de uma audiência negativa só se detecta em combinação com uma audiência positiva, porque aparece como redução da força do comportamento apropriado a esta; quando as duas estão juntas, instala-se um dilema, em que a resposta é emitida e punida por uma ou retida, com perda de prestígio perante a outra, e enquanto as audiências permanecem separadas o dilema não aparece. O orador sedicioso que vê um policial se aproximar fala menos à medida que a audiência negativa se torna relativamente mais importante, talvez até abaixo do nível aberto. O oposto é a audiência não punitiva: o psicoterapeuta precisa se estabelecer como tal, e, se o controle da audiência muda, o paciente pode emitir comportamento antes punido; o confidente é a pessoa simpática a quem se fala com menos medo de punição que a ouvintes ao acaso.

Dali se toma a definição usada no resto do capítulo. Ela é proposta aqui, e não de Skinner nem do Mackinac Center.

DEFINIÇÃO PROPOSTA

Janela de uma audiência.

Para um falante e um assunto, o conjunto de enunciados que esse falante emite na presença dessa audiência sem que ela os puna. “Punir” tem o sentido funcional de Skinner: é a consequência que torna o enunciado menos provável, o que inclui a perda de prestígio, e não só a sanção formal.

Duas consequências decorrem da definição. A primeira é que a janela é propriedade da relação entre falante, audiência e assunto, e não do enunciado: o mesmo enunciado está dentro da janela de uma audiência e fora da de outra, e não há “a” janela de uma sociedade, há tantas quanto as audiências relevantes. A janela de que a conversa política fala, a do “debate público”, é, na leitura aqui proposta, a das audiências que dominam uma arena num momento, e a análise precisa dizer quais. A segunda é que “dentro” e “fora” não são categorias do que é verdadeiro, justo ou correto: dizem o que costuma ou não costuma ser punido por quem ouve. Um enunciado falso pode estar dentro da janela de uma audiência, e um verdadeiro, fora da de outra. A definição não toma partido sobre nenhum dos dois.

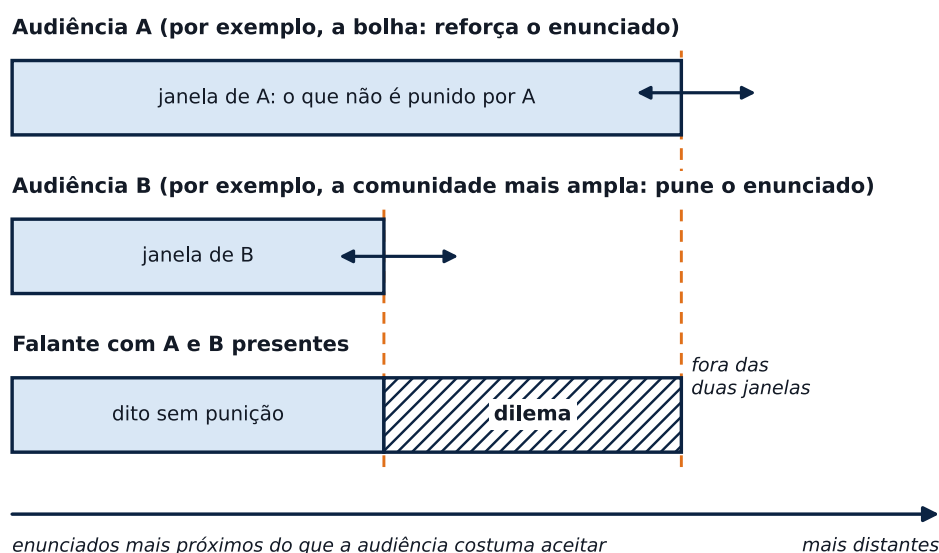
<p>■ DADO</p>	<p>O Mackinac Center descreve a janela como o conjunto de políticas que políticos podem apoiar sem perder apoio popular, diz que ela se move sobretudo pela lenta evolução de valores e normas e que políticos são mais seguidores que líderes; o próprio centro diz usar o conceito para promover suas políticas (Mackinac Center, 2019). Skinner (1957) descreve a audiência negativa, o dilema de audiências com janelas diferentes e a audiência não punitiva.</p>
<p>■ INTERPRETAÇÃO</p>	<p>A janela de uma audiência é o conjunto de enunciados que ela não pune, para um falante e um assunto; é relação, e não propriedade do enunciado.</p>
<p>■ HIPÓTESE</p>	<p>Políticos “seguem” na medida em que o que emitem é controlado pela punição das audiências que importam para eles. Testá-la exige medir a sanção, e não só o que foi dito; a página do Mackinac Center não diz como.</p>

10.2 Da janela da política à janela do enunciado: o que desloca a borda

A definição de janela como propriedade da relação entre falante, audiência e assunto (seção 10.1) tem uma implicação que a Figura 10.1 desenha. Duas audiências têm janelas de largura diferente para o mesmo assunto: a de A, por exemplo uma bolha (Capítulo 6) em que o enunciado é reforçado, é larga, e a de B, por exemplo uma comunidade mais ampla, é estreita. Quando só A está presente, o falante emite até a borda de A; quando só B está, até a de B. Quando as duas estão presentes, a faixa entre as bordas é a zona de dilema de Skinner: o enunciado é reforçado por A e punido por B, e o falante o emite e é punido por B ou o retém e perde prestígio diante de A. O que ele faz depende do peso relativo de cada audiência, que a figura não representa.

FIGURA 10.1

A janela é relativa à audiência: duas janelas e a zona de dilema



Esquema conceitual original, sem dados. Setas duplas: a borda se desloca (seção 10.2). Hachurado: reforçado por A e punido por B (Skinner, 1957, p. 230–231). Simplificação: as duas janelas supõem a mesma ordem de enunciados.

Nota. Esquema original, sem dados. As setas duplas indicam que a borda se desloca (ver o texto). A faixa hachurada é a zona de dilema (Skinner, 1957). Simplificação: as duas janelas supõem a mesma ordem de enunciados, do mais ao menos aceito; na realidade, duas audiências podem ordenar os enunciados de modos diferentes.

Em termos de contingências, a borda de uma janela se desloca quando muda a probabilidade de punição dos enunciados vizinhos a ela. Reúnem-se aqui sete candidatas a explicar a mudança, cada uma com o que as fontes verificadas fornecem. A primeira é quem faz parte da audiência: a entrada ou a saída de ouvintes com outro histórico muda o que é reforçado e o que é punido, e o Ca-

pítulo 6 tratou das bolhas como audiências que reforçam o que o grupo aceita. A segunda é o poder de sanção da audiência: Skinner (1957) fala em audiências negativas “efetivas”, e o contracontrole (Delprato, 2002), tratado na seção 6.4, é o conjunto de respostas que reduz a eficácia do controle aversivo; sem poder de sanção, a borda recua. A terceira é a informação sobre o que os outros toleram, o objeto das seções 10.3 e 10.4. A quarta é o que o ambiente mostra que os outros fazem, isto é, as normas descritivas, tratadas na seção 10.6. A quinta é o que recebe retorno na borda: enunciados limítrofes que são reforçados passam a ser emitidos com mais frequência, e a seção 10.5 trata disso com dados de políticos. A sexta é a repetição: encontrar repetidamente uma manchete falsa reduziu o quanto parecia antiético publicá-la e compartilhá-la, em quatro experimentos (Effron e Raj, 2020), resultado que o Capítulo 3 (seção 3.5) trata com mais vagar. A sétima é o alcance: a plataforma decide quantos ouvem cada emissão (Capítulo 8), e uma janela larga de uma audiência pequena e uma estreita de uma audiência enorme têm efeitos muito diferentes sobre o que circula.

■ DADO	Cada uma das variáveis acima tem apoio em ao menos um estudo verificado, de naturezas muito diferentes (experimentos, séries de tweets, um experimento com repetição de manchetes), e nenhuma foi manipulada junto com as outras.
■ INTERPRETAÇÃO	A mesma mudança de fala pode ser produzida por mais de uma delas, e o termo “deslocamento da janela” descreve o resultado, e não o mecanismo.
■ HIPÓTESE	As sete variáveis agem em conjunto na prática (por exemplo, a informação sobre a tolerância alheia chega por uma plataforma, que decide o alcance). Nenhuma fonte verificada separa as contribuições.

10.3 Informação sobre o que os outros toleram: o experimento de Pittsburgh

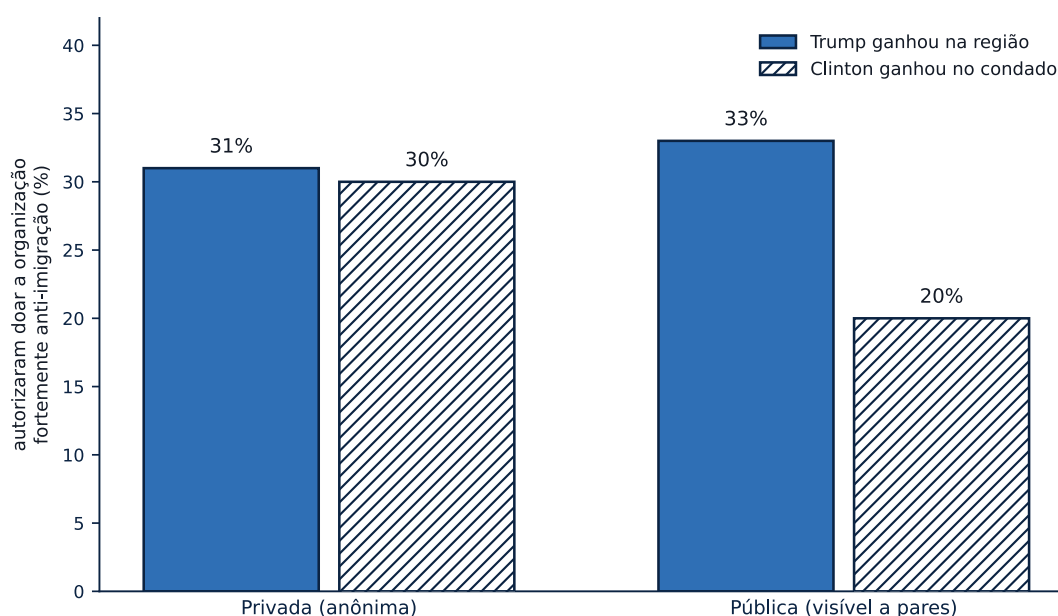
O experimento da abertura é o teste mais direto, entre as fontes verificadas, da hipótese de que informação sobre o que os outros aceitam desloca o que se faz em público. Bursztyn et al. (2020) o fizeram em dois experimentos que olham os dois lados da interação: quem age e quem observa e pode sancionar. No primeiro, de agosto a outubro de 2018, uma empresa de painel online recrutou 1.600 participantes dos sete condados da região metropolitana de Pittsburgh. Ao acaso, metade foi informada de que Trump ganhara na região metropolitana na eleição de 2016, e a outra metade, de que Clinton ganhara no condado de Allegheny, onde fica Pittsburgh. Ofereceu-se, então, um bônus em dinheiro a

quem autorizasse uma doação, em seu nome, a uma organização fortemente anti-imigração (para os autores, aceitar era uma ação xenófoba lucrativa), e metade dos participantes foi assegurada de anonimato (condição privada), enquanto a outra foi levada a crer que a decisão seria visível a pares da região (condição pública).

Em privado, as taxas de doação foram de 31% (Trump ganhou) e 30% (Clinton ganhou), sem diferença significativa ($p = 0,894$); em público, de 33% e 20% ($p < 0,001$). A diferença entre as diferenças, isto é, o quanto o efeito da informação em público excede o efeito em privado, foi significativa ($p = 0,008$). A percepção média da popularidade local das visões anti-imigração foi de 50,1% entre quem soube que Trump ganhara e de 42,6% entre quem soube que Clinton ganhara. Na estimação por variável instrumental, que usa o sorteio da informação para estimar o efeito da percepção sobre a doação, um ponto percentual a mais nessa percepção aumentou em 2,3 pontos percentuais a doação em público e não alterou a privada, resultado que os próprios autores tratam como sugestivo, porque o tratamento pode agir por outros canais, embora os julguem improváveis.

FIGURA 10.2

Doação a organização anti-imigração, em privado e em público, segundo quem ganhou a eleição na área



N = 1.600, região metropolitana de Pittsburgh. Diferença entre as diferenças: $p = 0,008$.

Nota. Valores arredondados, como no artigo (Bursztyn et al., 2020). N = 1.600, região metropolitana de Pittsburgh, 2018; a informação sobre o vencedor foi sorteada. O tratamento é a informação sobre quem ganhou a eleição na área, e não uma fala de candidato.

A Figura 10.2 mostra onde está a diferença, e é mais fina do que “a vitória de Trump liberou a xenofobia”. Os autores observam que o resultado vem sobretudo da cunha entre doar em privado e em público na condição Clinton (30% contra 20%), sem diferença significativa na condição Trump (31% e 33%; $p = 0,594$). Leem isso como indicação de que, em Pittsburgh, a vitória de Trump não gerou pressão significativa para pessoas tolerantes agirem como xenóforas em público, enquanto saber da vitória de Clinton se associou à pressão social para que xenófobos parecessem tolerantes em público. A leitura que fazem é a de que a informação não torna os participantes mais xenófobos: deixa mais à vontade, para expressar suas visões em público, os que já eram.

O segundo experimento olha o lado de quem observa e pode sancionar. Em novembro de 2018, cerca de 1.800 respondentes que se identificavam como democratas dividiram US\$ 2 com outra pessoa num jogo do ditador, em que quem decide quanto repassar fica com o resto. A outra pessoa era um participante do primeiro experimento que fizera a doação à organização anti-imigração (só entraram os que doaram), e o respondente não sabia que a área era Pittsburgh. Dois dados sobre esse doador foram cruzados ao acaso: quem ganhara a eleição na área dele, Trump ou Clinton, e se a doação fora privada ou visível a pares, o que dá quatro condições. O valor médio dado foi de US\$ 0,78, US\$ 0,81 e US\$ 0,78 nas três condições em que não havia razão para supor pressão por agir como xenófobo (as duas privadas e a pública onde Clinton ganhou) e de US\$ 0,88 na condição de doação pública em área onde Trump ganhou; a parcela que não deu nada foi de 18%, 21% e 22% nas três e de 8% na última. Os autores explicam o resultado pelo fato de que um doador público em área onde Trump venceu pode estar sob pressão social para agir como xenófobo, o que enfraquece a inferência de que seja realmente xenófobo; por isso recebe mais, isto é, é menos castigado.

Na linguagem aqui usada, o experimento 1 mexe na borda do lado do falante e o 2, no da sanção do ouvinte, e as duas bordas se movem com informação do mesmo tipo, a de quem ganhou a eleição na área. A informação não mudou o que os participantes aceitavam fazer onde ninguém veria, e mudou, ou deixou de estreitar, o que aceitavam fazer onde outros veriam. Isso é compatível com a definição da seção 10.1, em que a janela é propriedade da relação com uma audiência, e não do falante. Os autores adotam uma visão estreita de norma social, em que ela guia ações públicas ou potencialmente públicas e não privadas, e dizem que o que manipularam foi a popularidade percebida de Trump, e não falas dele.

Os limites são nítidos. A variável manipulada é a informação sobre quem ganhou a eleição na área, e não o que um político disse; o desfecho é autorizar uma doação, e não proferir um enunciado; a amostra é de um painel online de

uma região e, no experimento 2, só de democratas; e o caso é um só, de um lado da questão. O que o experimento sustenta é que a informação sobre o apoio local a uma posição altera o que as pessoas aceitam fazer em público quanto a ela, sem alterar o que fazem em privado. Que o mesmo valha para falas, para outras posições e para outras audiências é extensão aqui proposta.

■ DADO	Quem soube que Trump ganhara na área aceitou doar em público a uma organização anti-imigração em proporção semelhante à de quem doaria em privado (33% e 31%); quem soube que Clinton ganhara aceitou menos em público que em privado (20% e 30%). Observadores democratas deram menos a doadores públicos em área onde Clinton ganhou do que em área onde Trump ganhou (Bursztyn et al., 2020).
■ INTERPRETAÇÃO	Informação sobre a popularidade local de uma posição altera o custo percebido de expressá-la em público, de quem age e de quem sanciona, sem alterar a posição privada. Nos termos aqui usados: desloca-se a janela da audiência, não a disposição do falante.
■ HIPÓTESE	O mesmo vale para enunciados, para outras posições e para outras audiências. Os autores não manipulam falas.

10.4 Duas eleições, duas medidas: Brasil, 2018, e Estados Unidos, 2016

Se a informação sobre o que os outros toleram desloca o que se faz em público, uma eleição surpreendente é candidata natural a experimento da natureza: informa, de uma vez e a todos, que o apoio a um candidato e ao que ele diz é maior ou menor que o imaginado. Dois estudos verificados tentam ler isso em texto de rede social e chegam a resultados que não se reduzem um ao outro.

Marino-Fages e Martínez-Pozo (2026) estudam a eleição presidencial brasileira de 2018. Tratam o resultado como surpresa: Bolsonaro teve 46% dos votos no primeiro turno, bem acima das pesquisas, que previam cerca de 35%, e a surpresa variou entre municípios, porque o apoio a ele variou muito de um lugar para outro. Coletaram mais de 37,6 milhões de tweets em português, geolocalizados no Brasil, sem retuítes, publicados às segundas-feiras de julho de 2017 a dezembro de 2019, e atribuíram cada tweet a um município pela latitude e pela longitude. Para medir discurso de ódio usaram dois métodos. O primeiro foi um modelo de linguagem em português, o BERTimbau (uma versão do BERT), ajustado em 5.668 tweets de 2017 que outros pesquisadores haviam ro-

tulado como com ou sem discurso de ódio. O segundo foi um dicionário de termos que classifica o alvo do ódio em seis categorias (ódio político, alvo político, homofobia, racismo, sexismo e insultos). O desfecho em cada município é a parcela dos tweets classificados como ódio.

O desenho é de diferenças em diferenças: compara-se a mudança da parcela de ódio entre antes e depois da eleição, excluído o período de agosto a outubro de 2018, e vê-se como essa mudança varia com a parcela de votos de Bolsonaro no primeiro turno em cada município. A identificação supõe que, sem a eleição, as tendências teriam sido paralelas entre municípios com diferentes níveis de apoio a ele. No conjunto do país, a parcela média de tweets de ódio pelo BERT subiu de 8% (julho de 2017 a julho de 2018) para 9% (2019), cerca de 3.000 tweets a mais num dia médio.

O resultado central é que o discurso de ódio aumentou depois da eleição e aumentou mais onde Bolsonaro teve menos votos: em municípios com parcela de votos um desvio-padrão maior, o aumento foi 0,033 desvio-padrão menor com o classificador BERT e 0,018 com o dicionário (o desvio-padrão, medida da variação típica de cada variável, põe as duas na mesma escala). O aumento concentrou-se em homofobia e sexismo, temas da retórica do candidato; para insultos gerais, ódio político e alvos políticos, não houve padrão comparável, e os autores concluem que o resultado não é movido por polarização crescente. O número de usuários que passaram a postar conteúdo de ódio sem tê-lo feito antes (a margem extensiva) cresceu mais onde ele era menos popular, e a intensidade da expressão de quem já entrara (a margem intensiva) cresceu mais onde tinha mais apoio. A parcela de usuários que postam discurso de ódio subiu 10,1% entre homens, 6,1% entre mulheres e 3,3% entre contas pseudônimas ou não identificáveis. Como exemplo numérico, os autores tomam o Rio de Janeiro (58,3% dos votos para Bolsonaro; cerca de 32 mil tweets por dia útil, 9% com discurso de ódio): um município semelhante, com parcela de votos 10 pontos percentuais menor, geraria de 40 a 50 tweets de ódio a mais por dia, o que chamam de modesto no nível individual, mas cumulativo e de ampla participação.

A interpretação dos autores é que o padrão é compatível com uma quebra da espiral do silêncio, em que a percepção de ser minoria numa questão de valores leva ao autossilenciamento por medo de isolamento social: a eleição revelou que visões associadas à retórica do candidato eram mais difundidas do que se acreditava, reduzindo o custo social percebido de expressá-las, sobretudo onde ele era menos popular. É, no vocabulário aqui usado, a terceira candidata da seção 10.2, e coincide com a leitura de Bursztyn et al. (2020). Em favor do desenho, os autores observam que o Brasil não tinha, no período, marco legislativo abrangente sobre moderação de conteúdo em plataformas, o que tor-

na improvável que novos regimes legais expliquem o aumento. Cabem quatro ressalvas. O estudo é documento de trabalho, e a própria capa diz que não passou por revisão de pares. O classificador teve acurácia, precisão, revocação e F1 de 77 a 79% nos conjuntos de validação e de teste (acurácia é a proporção de tweets classificados corretamente; precisão, a de acertos entre os marcados como ódio; revocação, a dos tweets de ódio que ele achou; o F1 combina as duas últimas). Os tweets de segunda-feira e os geolocalizados são tratados como amostras representativas do universo de tweets, pressuposto dos autores, e a razão de geolocalizados para o total varia de 4% a 8% por palavra. E o desenho mede texto, e não percepção de custo social: o mecanismo é leitura dos autores, e não resultado.

Siegel et al. (2021) estudam a eleição americana de 2016. Reuniram dois conjuntos de tweets, de 17 de junho de 2015, o dia seguinte ao do anúncio da candidatura de Trump, a 15 de junho de 2017: todos os tweets que citavam Trump ou Clinton (mais de 600 milhões sobre ele e mais de 150 milhões sobre ela), coletados em tempo real, de modo que tweets depois apagados ou de contas depois suspensas continuam nas contagens; e quase 400 milhões de tweets de uma amostra aleatória de 500 mil usuários americanos, para ver se o ódio também aumentou fora da conversa política.

Definiram discurso de ódio como linguagem hostil e maliciosa, motivada por viés, dirigida a uma pessoa ou grupo por características inatas, reais ou percebidas; elogio à Ku Klux Klan ou orgulho excessivo do próprio grupo, em si, não contam, e a retórica nacionalista branca foi medida à parte. Usaram dois métodos. O primeiro foi um dicionário de 3.268 termos (insultos contra asiáticos, negros, imigrantes, muçulmanos, judeus, homossexuais e mulheres, mais a retórica nacionalista branca), tirado de bases públicas. Como muitos termos apareciam em nomes de usuário, dentro de outras palavras ou em tweets que condenavam o ódio, cerca de 25.000 tweets com esses termos foram classificados à mão, por três avaliadores cada um, e serviram para treinar classificadores (Naive Bayes) que eliminam esses falsos positivos; o de discurso de ódio teve 94% de acurácia, 95% de precisão e 90% de revocação. O segundo método dispensa saber as palavras de antemão: mede o quanto os tweets de cada dia se parecem com comentários de comunidades do Reddit conhecidas por conteúdo racista e de direita alternativa, tomados como discurso de ódio “na natureza”, e só comparações relativas, como a dos tweets depois da eleição com os do início da campanha, são interpretáveis.

A análise é de série temporal interrompida, que compara o nível diário antes e depois de um evento, e os autores calcularam quatro medidas (número e proporção de tweets e de usuários com ódio por dia), que contam essencialmente a mesma história. Em qualquer mês, entre 0,1% e 0,3% dos tweets continham

ódio. Não observaram aumento persistente de discurso de ódio ou de linguagem nacionalista branca nem ao longo da campanha nem depois da eleição de Trump; eventos-chave geraram picos breves, que se dissiparam rapidamente, e o maior pico mensal nos dados sobre Trump, no fim de janeiro de 2017, foi em grande parte de ódio misógino depois do decreto de restrição de viagens. Dizem que o dicionário é conservador, porque exige que o tweet contenha insulto ou termo depreciativo. Limitam o escopo ao Twitter, observam que o discurso de ódio pode ter proliferado em plataformas especializadas e deixam para pesquisa futura o efeito do banimento de contas, e advertem que ver certo discurso num momento não significa que ele tenha aumentado, dado o caráter em rajadas do Twitter, em que tópicos ganham destaque por pouco tempo e voltam ao equilíbrio.

Os dois estudos não se contradizem de modo simples, e também não se somam. Diferem na medida (um classificador treinado em tweets rotulados e um dicionário por alvo, contra um dicionário que exige insulto e uma classificação por comunidades), no desenho (diferenças em diferenças entre municípios, contra séries temporais da campanha e da eleição), no país, na plataforma e na pergunta: o aumento diferencial segundo a popularidade local do candidato, de um lado, o nível agregado do discurso no Twitter americano, de outro. Um contraste local pode detectar o que uma série agregada não detecta, e o contrário também ocorre; a que se deve a diferença é o que os dados não decidem. A discordância é parte do dado, e o capítulo não a resolve.

<p>■ DADO</p>	<p>No Brasil, o discurso de ódio nos tweets geolocalizados aumentou depois da eleição de 2018, mais onde Bolsonaro teve menos votos, concentrado em homofobia e sexismo (Marino-Fages e Martínez-Pozo, 2026, documento de trabalho). Nos Estados Unidos, não houve aumento persistente de discurso de ódio no Twitter durante a campanha de 2016 nem depois da eleição, com a medida adotada (Siegel et al., 2021).</p>
<p>■ INTERPRETAÇÃO</p>	<p>Os resultados são compatíveis com a ideia de que uma eleição informa sobre a tolerância alheia, e discrepam quanto ao efeito agregado; as medidas, os desenhos e os contextos diferem demais para que um confirme ou refute o outro.</p>
<p>■ HIPÓTESE</p>	<p>A informação trazida pela eleição reduziu a sanção percebida por falantes que antes se continham. Nenhum dos dois estudos mede a sanção percebida nem a que de fato ocorreu.</p>

10.5 Políticos e redes sociais: quem fala de cima e o que recebe

Se, como diz o Mackinac Center (seção 10.1), políticos são mais seguidores que líderes, resta saber quem eles seguem. Frimer et al. (2023) oferecem o estudo verificado que mais se aproxima de responder, para políticos americanos no Twitter, e ele fala menos de “o público” e mais do retorno que a plataforma devolve. Aplicando um classificador de incivilidade a 1,3 milhão de tweets de membros do Congresso desde 2009, os autores observam aumento de 23% na incivilidade ao longo de uma década (de 11,8 em 2009 a 14,5 em 2019). O classificador, a PerspectiveAPI, foi feito para auxiliar a moderação de fóruns e dá a cada texto uma nota de 0 a 100, que reflete a probabilidade de um observador achá-lo rude ou desrespeitoso. O aumento se manteve nos modelos que dão a cada político o seu próprio nível de base (intercepto por político), de modo que é mais atribuível a políticos que se tornaram mais incivis do que à entrada de políticos mais incivis. Numa amostra pseudoaleatória de 998.935 tweets dos Estados Unidos, a incivilidade também subiu entre civis (de 21,1 a 27,0, 28%), mas a alta entre os políticos persistiu com esse controle; subiu no Senado e na Câmara e em todo o espectro político, com a maior elevação entre democratas liberais, diferença que os autores explicam sobretudo por reações ao presidente Trump. Nos tuítes dos presidentes, a incivilidade média foi de 13,4 com Obama, sem mudança no tempo, e subiu a 18,8 em 2017 e a 23,0 em 2019 com Trump.

A parte que interessa à análise operante é a do retorno. Tuítes extremamente incivis receberam em média dez vezes mais retuítos e oito vezes mais curtidas que os extremamente civis, e a razão cresceu: em 2009, os incivis recebiam cerca de duas vezes mais retuítos; em 2019, quase 15 vezes (curtidas: 1,3 vez e quase 8 vezes). A análise de mediação pergunta se o retorno explica o aumento: ao controlar estatisticamente o retorno em curtidas, o aumento da incivilidade caiu 74%; ao controlar o de retuítos, desapareceu. Como a mediação é só correlacional, os autores a completaram com uma análise de causalidade de Granger dentro de cada conta, que compara os momentos em que a incivilidade foi especialmente bem recebida com a incivilidade que veio depois, e a análise mostrou que períodos de melhor recepção foram seguidos de mais incivilidade. Concluem que os políticos aprenderam com a experiência positiva de tuitar com incivilidade e, por isso, intensificaram-na, e atribuem o aumento tanto à substituição, no Congresso, de políticos menos incivis por mais incivis quanto à mudança dos próprios políticos.

Dois resultados do mesmo estudo ensinam a cautela. O primeiro: nos discursos no plenário do Congresso (1995 a 2020) e nos debates presidenciais (1976 a 2016), a incivilidade não subiu e pode ter caído, o que os autores tomam como apoio à ideia de que as redes sociais podem fomentá-la. O segundo: curtidas e retuítes não implicam aprovação. Em estudos do suplemento, quanto mais curtidas um tuíte recebeu, menos as pessoas de fato gostaram dele, e os retuítes de tuítes incivis parecem movidos pelo valor de entretenimento, e não por aprovação moral; a contagem absoluta de curtidas é confundida pela maior exposição dos tuítes incivis, e para corrigir isso os autores criaram a razão curtidas por retuíte. Registram ainda que a amostra reúne 841 dos 1.003 membros do Congresso das legislaturas 111 a 116, que a API limita a coleta a 3.240 tweets por conta, e que a concordância do classificador com observadores humanos foi $r = 0,627$.

Lido com o vocabulário do capítulo, o estudo mostra três coisas. A primeira é que o que o político emite no Twitter acompanhou o retorno que a plataforma devolve, e que os enunciados na borda (os incivis) receberam mais retorno que os civis. A segunda é que a audiência a que o político “segue” não é “o público”, e sim o subconjunto que reage, cuja reação, segundo os próprios autores, diverge da aprovação. A terceira é que a incivilidade não subiu nas arenas presenciais (plenário, debates), o que é compatível com a ideia de que a janela é de cada arena, e não do político, como a definição da seção 10.1 supõe. O Capítulo 9 (seção 9.6) já viu, em congressistas americanos, que as publicações sobre o exogrupo foram mais compartilhadas nos dois campos (Rathje et al., 2021); o estudo de Frimer et al. (2023) mostra aumento da incivilidade em todo o espectro político, com a maior elevação entre democratas liberais.

O estudo não responde se os políticos moveram a janela da audiência. Mede o comportamento do político e o retorno que recebe, e não o que os eleitores passaram a dizer. Os três desenhos reunidos até aqui cobrem cada um um lado: Bursztyn et al. (2020) manipulam a informação que as pessoas têm sobre o apoio local; Marino-Fages e Martínez-Pozo (2026) observam a fala de usuários depois de uma eleição; Frimer et al. (2023) observam a fala de políticos e o retorno. Nenhum liga a fala do político à da audiência com um desenho que identifique a direção. Perguntar se o político lidera ou segue, tal como a página do Mackinac Center a formula, é perguntar o que nenhuma das três fontes mede.

Há, por fim, uma observação de método que serve de aviso. Kahmann e Heyer (2019) propõem um método de processamento de linguagem natural para achar, em textos de jornal, enunciados que deslocam as bordas da janela. A ideia é simples: num período de referência (8.067 documentos do jornal *taz* sobre refugiados, de 2013 e 2014), mapeiam-se as palavras que costumam apa-

recer junto às do tema, o contexto-padrão; depois, cada frase de 2015 recebe uma distância, de 0 a 1, que cresce com a proporção de palavras que nunca apareceram nesse contexto, e só se mantêm as frases com indicador de enunciado, como verbos de dizer e aspas, o que reduz 14.894 frases com palavra-alvo a 3.126. Os próprios autores reconhecem que o uso de palavras novas nem sempre reflete uma opinião fora da janela, porque muitas co-ocorrências recebem distância alta só por não terem sido usadas antes, e esse não uso se deve a efeitos aleatórios, e não necessariamente a ser politicamente incorreto. Reconhecem também que o método não garante que toda frase achada seja extrema: entre as de maior distância aparece uma sobre o sucesso educacional de migrantes, que não parece exceder nenhuma janela (leitura aqui proposta). O que o método mede é novidade de contexto, e não sanção; confundir as duas é o erro que o capítulo quer evitar.

■ DADO	Em 1,3 milhão de tweets de congressistas americanos, a incivilidade subiu 23% em uma década, com retorno (retuítes e curtidas) muito maior para tuítes incivis, controle do retorno reduzindo ou eliminando o aumento, mediação correlacional e análise de Granger a favor do reforçamento, parte do aumento por substituição de políticos, e sem aumento nos discursos de plenário (Frimer et al., 2023).
■ INTERPRETAÇÃO	O que políticos emitem na rede varia com o retorno que a plataforma devolve, e enunciados na borda foram reforçados; a arena importa, porque o mesmo político não aumentou a incivilidade no plenário.
■ HIPÓTESE	Políticos moveram a janela das suas audiências, ou as seguiram. Nenhuma fonte verificada mede a fala das audiências em relação à dos políticos com desenho que identifique a direção.

10.6 Sanção, contracontrole e o que mantém a borda no lugar

O Capítulo 5 tratou do discurso de ódio como mando disfarçado e como controle aversivo; este capítulo toma dele só o que serve à janela: o que faz a borda recuar. O estudo verificado mais claro sobre isso é o experimento de Álvarez-Benjumea e Winter (2018). Participantes de uma plataforma de crowdsourcing (180, recrutados na Clickworker, residentes na Alemanha e pagos com três euros) comentaram, num fórum montado pelos autores, nove fotos de quatro temas controversos (feminismo, direitos LGBT, refugiados e multiculturalismo, e pobreza). Eram obrigados a comentar cada discussão e viam apenas comentários escolhidos pelos autores, tirados de uma sessão prévia em que três avaliadores externos classificaram 840 comentários como amisto-

sos, neutros ou hostis. As discussões variavam em quatro condições: base (dois comentários amistosos, dois neutros e dois hostis), censurada (o conteúdo de ódio anterior apagado), extremamente censurada (só comentários amistosos) e contra-fala (os comentários hostis acompanhados de réplicas que destacam a inaceitabilidade do ódio).

Os comentários dos participantes foram avaliados por três juízes cegos às condições, numa escala de 1 (muito amistoso) a 9 (muito hostil), com concordância relativamente alta entre eles (alfa de Krippendorff de 0,69, índice de concordância em que 1 é concordância perfeita); a média das três notas formou o escore de hostilidade. Comentários hostis (nota 7 ou mais) foram 24 (5,88%) na base, 4 (0,99%) na condição censurada, 18 (4,44%) na contra-fala e 15 (4,10%) na extremamente censurada; só a redução da condição censurada foi significativa. A contra-fala não teve efeito significativo, e os autores sugerem como explicação possível que os contracomentários, sendo eles próprios hostis, podem acrescentar ambiguidade sobre a norma ao pôr em conflito a informação descritiva e a injuntiva. Ressalvam que não têm dados sobre a troca entre censura e liberdade de expressão e que o anonimato do fórum impede processos de identificação de grupo; o estudo não toma posição sobre censurar. O Capítulo 5 (seção 5.5) retoma, com os mesmos dados, a conclusão dos autores de que o comportamento normativo online é motivado mais por normas descritivas (o que os outros fazem) do que por normas injuntivas (o que se deve fazer).

Em termos de Skinner (1957), a contra-fala é uma sanção sem poder: partiu de pares anônimos, sem controle sobre reforçadores do falante, e nada no desenho a transforma em audiência negativa “efetiva”. Longe de ser um achado definitivo, trata-se de uma leitura analítica derivada da explicação dos autores (segundo a qual a própria réplica carrega hostilidade), já que a pesquisa não manipulou a autoridade da fonte sancionadora. O resultado importa, porém, para a prática: nos dados do experimento, a hostilidade não diminuiu porque alguém disse que não devia, diminuiu quando o ambiente exibiu menos dela. A mesma regra deveria valer no sentido contrário, de modo que um ambiente que passa a exibir mais o que antes estava fora da janela a alargaria; é hipótese aqui proposta, e o experimento não a testou.

O contracontrole (Delprato, 2002) entra pelo mesmo caminho. A sanção que pune um enunciado é uma contingência aversiva socialmente mediada; quem a sofre pode fugir ou se esquivar dela, e um dos quatro repertórios de esquiva descritos na seção 6.4 é o do lugar, a migração para espaços fechados e sem moderação. Em termos de Skinner (1957), é buscar o confidente, a audiência não punitiva. Daí a hipótese, aqui proposta: a bolha do Capítulo 6 é uma audiência não punitiva coletiva para certos assuntos, que alarga a janela efetiva do falante enquanto ele permanece nela, e o dilema de Skinner aparece quan-

do uma publicação alcança, ao mesmo tempo, a bolha e uma audiência com janela mais estreita. Skinner (1957) diz que, enquanto as audiências permanecem separadas, o dilema não aparece. Se as plataformas juntam numa mesma emissão audiências antes separadas, o dilema se instala com mais frequência; nenhuma fonte verificada mede isso.

TABELA 10.1

O que cada estudo manipulou ou observou, o que mediu como “dizer” e o que não permite concluir

Estudo	Material e desenho	O que mediu como “dizer”	O que observaram	Ressalva
Bursztyn et al. (2020)	Dois experimentos online nos EUA, 2018: 1.600 moradores de Pittsburgh; cerca de 1.800 democratas. Informação sorteada sobre quem ganhou a eleição na área	Autorizar doação a organização anti-imigração, em privado e em público (ação); no exp. 2, valor dado ao doador	Em público, 33% (Trump ganhou) e 20% (Clinton ganhou); em privado, 31% e 30%. Observadores deram mais a doador público onde Trump ganhou	Manipula informação, não fala de político; ação, não enunciado; painel online; observadores só democratas
Marino-Fages e Martínez-Pozo (2026)	Eleição de 2018 no Brasil como surpresa; diferenças em diferenças por município; mais de 37,6 milhões de tweets geolocalizados	Discurso de ódio por classificador BERT (F1 de 77 a 79%) e por dicionário	Aumento depois da eleição, maior onde Bolsonaro teve menos votos; concentrado em homofobia e sexismo	Documento de trabalho, sem revisão de pares; só 4 a 8% geolocalizados; tendências paralelas supostas
Siegel et al. (2021)	Séries de mais de 750 milhões de tweets da eleição de 2016 e de quase 400 milhões de amostra aleatória	Dicionário aumentado por aprendizado de máquina e classificação por comunidades do Reddit	Sem aumento persistente de discurso de ódio na campanha nem depois da eleição; picos breves	Dicionário exige insulto; só Twitter; autores indicam outras plataformas
Frimer et al. (2023)	1,3 milhão de tweets de congressistas americanos, 2009–2019; discursos de plenário e debates	Incivildade por classificador (PerspectiveAPI; $r = 0,627$ com observadores)	+23% em uma década; retorno maior para tuítes incivis; sem aumento no plenário	Manuscrito aceito; incivildade não é discurso de ódio; mediação correlacional; parte por substituição
Álvarez-Benjumea e Winter (2018)	Experimento online com 180 participantes de crowdsourcing; fórum com quatro condições	Comentários com nota de hostilidade igual ou maior que 7	5,88% na base; 0,99% censurada; 4,44% contra-fala; 4,10% extremamente censurada	Participantes pagos e anônimos; sem dados sobre a troca com liberdade de expressão

Estudo	Material e desenho	O que mediu como “dizer”	O que observaram	Ressalva
Kahmann e Heyer (2019)	Método aplicado a jornal alemão (taz); 8.067 documentos de referência (2013 e 2014) e textos de 2015	Novidade de contexto de cada frase	Lista de frases de maior distância	Novidade não é impermissibilidade; os autores admitem que nem toda frase achada é extrema

Nota. Os números vêm dos artigos. As páginas de Bursztyn et al. (2020) são as do American Economic Review e as de Kahmann e Heyer (2019), as dos anais ; as demais fontes não têm paginação de revista nas versões lidas. A coluna “O que mediu como ‘dizer’” descreve o que cada estudo contou; nenhum mediu a sanção que o dizer recebeu, e só o de Álvarez-Benjumea e Winter (2018) a variou.

■ DADO	Num fórum experimental, remover o conteúdo de ódio anterior reduziu os comentários hostis de 5,88% para 0,99%, e acrescentar réplicas que apontam a inaceitabilidade não teve efeito significativo (Álvarez-Benjumea e Winter, 2018).
■ INTERPRETAÇÃO	O que o ambiente exhibe como feito pelos outros pesa mais, nesses dados, que a réplica verbal; a réplica de pares sem poder não tem o estatuto de audiência negativa efetiva (leitura aqui proposta, com base em Skinner, 1957).
■ HIPÓTESE	A bolha é audiência não punitiva coletiva para certos assuntos, e a plataforma, ao juntar audiências, multiplica os dilemas. Nenhum estudo verificado as testa.

10.7 Como situar uma fala em relação à janela

O capítulo não oferece um critério para dizer que uma fala está “fora da janela”: isso exigiria medir a sanção, e as fontes não a medem. O que se pode oferecer é um conjunto de perguntas que impede os erros mais comuns da conversa corrente. O instrumento é para o analista, e não classifica falas.

As perguntas 1 a 4 vêm das definições das seções 10.1 e 10.2; as 5 a 8, dos limites dos estudos reunidos na Tabela 10.1. A pergunta 7 repete o critério do livro inteiro: o mesmo método para todos os campos.

Perguntas para situar uma fala em relação à janela

1. Qual é a audiência e em que arena a fala ocorreu (plenário, rede, rádio, família)? A mesma pessoa pode ter janelas diferentes em arenas diferentes.
2. O que se sabe sobre a sanção que enunciados semelhantes receberam dessa audiência, por registro (e não por impressão)? Sem registro, o analista deve dizer que não sabe.
3. Há audiências com janelas diferentes presentes na mesma emissão (dilema)? A plataforma decide quais.
4. O que mudou entre dois momentos: quem escuta, o poder de sanção, a informação sobre a tolerância alheia, o que o ambiente exhibe, o que é reforçado na borda, a repetição ou o alcance?
5. O que se mediu como “dizer”: texto bruto, classificador, dicionário, ação? Com que erro e com que definição de “ódio” ou “incivilidade”?
6. Contra o que se comparou: o mesmo falante antes e depois, ou outros falantes, e que critério de seleção?
7. O mesmo critério foi aplicado a falas de campos políticos opostos?
8. O que a medida não distingue (novidade, incivilidade, ódio, ação)?

10.8 Interpretações rivais e limites

A leitura do capítulo, de que a janela de uma audiência é o conjunto de enunciados que ela não pune e de que a borda se desloca quando muda a sanção, tem rivais. O quadro abaixo as reúne, uma a uma, e os limites do capítulo vêm depois dele.

Interpretações rivais para o mesmo conjunto de dados

Mudança de composição, e não de janela. Frimer et al. (2023) atribuem o aumento tanto à substituição de políticos menos incivis por mais incivis quanto à mudança dos próprios políticos; o que se vê no agregado é a soma das duas, e só a segunda é mudança de fala de quem já falava.

Mudança de medida. Os estudos usam classificadores de F1 de 77 a 79% (Marino-Fages e Martínez-Pozo, 2026), dicionários que exigem insulto (Siegel et al., 2021) e um classificador de incivilidade com $r = 0,627$ com observadores (Frimer et al., 2023); “discurso de ódio” e “incivilidade” não são o mesmo constructo.

Outros canais, e não a norma. Bursztyn et al. (2020) dizem que o efeito do tratamento pode agir por outros canais, embora os julguem improváveis; a leitura em termos de norma é dos autores.

Evento, e não janela. Marino-Fages e Martínez-Pozo (2026) excluem os meses de agosto a outubro de 2018 e supõem que, sem a eleição, as tendências seriam paralelas entre municípios; a suposição pode falhar, e outros eventos do período podem ter agido diferencialmente entre municípios (observação aqui feita).

Metáfora que não se refuta. Se qualquer mudança de fala vale como deslocamento da janela, o conceito não separa hipóteses. Por isso aqui ele é definido pela sanção da audiência (seção 10.1), e a medida da sanção é pedida (seção 10.7).

Direção e simetria. Os estudos de eleição tratam das vitórias de Trump e de Bolsonaro; nenhum trata do caso espelhado. O conceito é simétrico por definição, e o dado mais próximo de simetria, no sentido de que o aumento foi geral, vem de Frimer et al. (2023), em que a maior elevação ocorreu entre democratas liberais, explicada pelos autores sobretudo por reações a Trump; a simetria do efeito de uma eleição não foi testada.

Fonte interessada. A descrição do conceito vem de um centro que o usa para promover as próprias políticas (Mackinac Center, 2019).

Amostra e plataforma. Painel online de uma região e observadores democratas (Bursztyn et al., 2020), participantes pagos de crowdsourcing (Álvarez-Benjumea e Winter, 2018), Twitter (os demais). Nada aqui diz de WhatsApp, YouTube ou TikTok.

Cinco limites valem para o capítulo inteiro. Primeiro, o conceito de janela, na origem, é sobre política pública e vem de parte interessada; a tradução para enunciados e audiências é proposta aqui, e nenhuma fonte a testa como tal. Segundo, as medidas são diferentes (percepção, ação pública, texto classificado) e não se somam; as comparações da Tabela 10.1 são de ordem de grandeza e de desenho, e não de efeitos. Terceiro, nenhuma fonte mede a sanção, que é o termo central da definição: o experimento de Bursztyn et al. (2020) toca o lado do observador, e o de Álvarez-Benjumea e Winter (2018) varia o que o ambiente exhibe, mas nenhum mede a punição que o falante de fato sofre. Quarto, a fonte brasileira é um documento de trabalho sobre uma plataforma e uma eleição, e nenhuma mede a fala de políticos brasileiros; essa é a pergunta que se quer fazer, e as fontes verificadas não a respondem. Quinto, os dados que ligariam a fala do político à da audiência, com identificação da direção, provavelmente exigem registros que só as plataformas têm, como nos limites do Capítulo 8.

Para quem quer usar o capítulo na análise de falas (Capítulos 3, 5 e 7), a consequência é prática. Dizer que uma fala “deslocou a janela” é afirmar três coisas, todas empíricas: que a audiência a puniria antes, que deixou de puni-la e que a fala do político teve parte nisso. O analista que afirma só a primeira coisa descreve uma impressão. O Capítulo 11 retoma, ao tratar das medidas de difusão, a distinção entre alcance, exposição e reação, que importa aqui porque a

janela efetiva de uma publicação depende de quem a vê; o Capítulo 12 (seção 12.6) trata de intervenções sobre as contingências da plataforma, e o critério daqui, o mesmo para o que alarga e para o que estreita a janela, vale para elas.

Síntese

A janela de Overton, na origem, é um modelo sobre quais políticas um político pode apoiar sem perder apoio popular, formulado por um centro que o usa para promover as próprias políticas e que afirma que os políticos são mais seguidores que líderes. Aqui ela é traduzida para a análise do comportamento verbal: a janela de uma audiência, para um falante e um assunto, é o conjunto de enunciados que essa audiência não pune; é relação, e não propriedade do enunciado; há tantas janelas quanto audiências, e o dilema de Skinner aparece quando duas audiências com janelas diferentes estão presentes (seções 10.1 e 10.2).

Os dados verificados mostram pedaços diferentes do fenômeno. Informação sobre quem ganhou uma eleição local alterou o que moradores de Pittsburgh aceitaram fazer em público, sem alterar o que fizeram em privado, e alterou o que observadores deram a quem agiu em público (10.3). Uma eleição brasileira surpreendente foi seguida de aumento diferencial de discurso de ódio em tweets, maior onde o candidato foi menos votado, num documento de trabalho; uma série americana não achou aumento persistente no Twitter, com outra medida (10.4). Congressistas americanos ficaram mais incivis no Twitter na medida em que a incivilidade recebia mais retorno, e não no plenário, com parte do aumento por substituição de políticos e com a ressalva de que curtidas não são aprovação (10.5). Num fórum experimental, remover o ódio reduziu o ódio, e a réplica que o condena não (10.6).

O que os dados não mostram é o termo central e a direção que a pergunta de fundo pede. Nenhum estudo mede a sanção, nem liga a fala do político à da audiência com identificação de direção, e nenhum trata de um caso espelhado. A janela é, nas fontes, o nome de um resultado (a mudança no que se faz ou diz em público) e não de um mecanismo; o capítulo oferece, na seção 10.7, perguntas para o analista distinguir o que sabe do que supõe. Os Capítulos 3, 5 e 7 usam a ideia de que a janela é por audiência e por arena; o Capítulo 6, a de que a bolha é audiência não punitiva (hipótese); o Capítulo 8, a de que o alcance decide quantos ouvem; o Capítulo 9, a de que o político é produtor que responde ao retorno; o Capítulo 12, a de que intervir é mexer nas contingências que mantêm a borda.

Referências

- Álvarez-Benjumea, A., & Winter, F. (2018). Normative change and culture of hate: An experiment in online environments. *European Sociological Review*, 34(3), 223–237. <https://doi.org/10.1093/esr/jcy005>
- Bursztyn, L., Egorov, G., & Fiorin, S. (2020). From extreme to mainstream: The erosion of social norms. *American Economic Review*, 110(11), 3522–3548. <https://doi.org/10.1257/aer.20171175>
- Delprato, D. J. (2002). Countercontrol in behavior analysis. *The Behavior Analyst*, 25(2), 191–200. <https://doi.org/10.1007/bf03392057>
- Effron, D. A., & Raj, M. (2020). Misinformation and morality: Encountering fake-news headlines makes them seem less unethical to publish and share. *Psychological Science*, 31(1), 75–87. <https://doi.org/10.1177/0956797619887896>
- Frimer, J. A., Aujla, H., Feinberg, M., Skitka, L. J., Aquino, K., Eichstaedt, J. C., & Willer, R. (2023). Incivility is rising among American politicians on Twitter. *Social Psychological and Personality Science*, 14(2), 259–269. <https://doi.org/10.1177/19485506221083811>
- Kahmann, C., & Heyer, G. (2019). Measuring context change to detect statements violating the Overton window. *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 392–396. <https://doi.org/10.5220/0008191803920396>
- Mackinac Center for Public Policy. (2019). *The Overton window*. <https://www.mackinac.org/OvertonWindow>
- Marino-Fages, D., & Martínez-Pozo, A. (2026). *Hate in the Tropics: Political leaders and the social acceptability of online hate speech* (Documento de trabajo N.º 391). RedNIE. <https://rednie.eco.unc.edu.ar/files/DT/391.pdf>
- Rathje, S., Van Bavel, J. J., & van der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, 118(26), Article e2024292118. <https://doi.org/10.1073/pnas.2024292118>
- Siegel, A. A., Nikitin, E., Barberá, P., Sterling, J., Pullen, B., Bonneau, R., Nagler, J., & Tucker, J. A. (2021). Trumping hate on Twitter? Online hate speech in the 2016 U.S. election campaign and its aftermath. *Quarterly Journal of Political Science*, 16(1), 71–104. <https://doi.org/10.1561/100.00019045>
- Skinner, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts.



RÁDIO IW4

Artigos, palestras e aulas para ouvir

Leituras de artigos científicos, palestras e aulas em áudio, 24 horas por dia, 7 dias por semana: para estudar no caminho, no trânsito ou na academia.

24h por dia, 7 dias por semana

Acesse, de graça e sem cadastro

walden4.com.br/radio



11

Infodemiologia: como a informação se espalha e o que se mede

Cinelli et al. (2020), num estudo sobre a infodemia de Covid-19, abrem o artigo com um caso. Segundo eles, uma emissora de notícias antecipou o rumor de que a Lombardia, no norte da Itália, seria fechada e o publicou horas antes da comunicação oficial do primeiro-ministro; em consequência, pessoas lotaram trens e aeroportos para ir ao sul antes do fechamento, o que, nas palavras dos autores, atrapalhou a iniciativa do governo de conter a epidemia e possivelmente aumentou o contágio. O caso é relatado como os autores o contam; a notícia não foi lida, e não se sabe como o comportamento das pessoas foi medido.

O caso serve ao capítulo por uma razão que não é a dos autores. A informação que circulou não é apresentada como falsa: era um rumor sobre uma medida que veio a ser tomada, e o dano que lhe atribuem vem de quando e de como as pessoas agiram sobre ela. A saúde pública distingue o rumor da desinformação por isso mesmo, porque o rumor pode vir a ser verdadeiro (Calleja et al., 2021). A leitura é aqui proposta, e a primeira lição do capítulo sai dela: o que preocupa quem estuda uma infodemia não é só a informação falsa, é o volume, a velocidade e a qualidade da informação, e a veracidade de cada item é uma variável entre outras importantes de serem consideradas.

O capítulo pergunta como se mede a difusão de informação, o que os dados verificados mostram sobre o verdadeiro e o falso, e como isso se liga às contingências dos capítulos anteriores. A seção 11.1 apresenta a infodemiologia e a palavra infodemia; a 11.2 trata das medidas de difusão e da distinção entre alcance, exposição e reação; a 11.3 examina o resultado de que o falso se espalha mais rápido que o verdadeiro; a 11.4, os modelos epidemiológicos aplicados à

informação; a 11.5, a infodemia de Covid-19, no mundo e no Brasil; a 11.6, o que o campo propõe fazer; e a 11.7 reúne perguntas para ler um estudo de difusão.

Três cautelas orientam a leitura. A primeira é que a linguagem epidemiológica é analogia. Nenhum agente infeccioso passa de uma pessoa a outra; o que se espalha são enunciados, e cada passagem é uma resposta verbal de alguém (Capítulo 3), mantida por consequências que o modelo de contágio não tem como representar. A analogia é útil para medir e arriscada para explicar. A segunda é que as medidas de difusão contam respostas de repassar, e não crença, nem efeito sobre o que as pessoas fazem depois. A terceira é a cobertura: quase tudo o que há vem do Twitter, em inglês, e do começo da pandemia; para o Brasil, o capítulo tem um único estudo de links sobre vacinas.

11.1 Infodemiologia e infovigilância

O termo infodemiologia é de Eysenbach, que o define como a ciência da distribuição e dos determinantes da informação num meio eletrônico, com o fim de informar a saúde pública e as políticas públicas (Eysenbach, 2009). O nome junta informação e epidemiologia, e a analogia é literal na proposta: como a epidemiologia conta casos e procura determinantes, a infodemiologia conta ocorrências de informação e procura o que as determina. O artigo de 2009 distingue dois tipos de indicador: os da oferta, que contam o que é publicado (páginas, mensagens, blogues), e os da demanda, que contam o que as pessoas procuram e por onde navegam. As métricas básicas são a prevalência de informação, o número absoluto ou relativo de ocorrências de uma palavra-chave ou conceito num conjunto; a incidência, que é quantas unidades novas de informação surgem por unidade de tempo; e a razão de ocorrência, que compara um conceito com um termo de controle, porque o total do conjunto costuma não ser conhecido. Usar esses dados para vigilância é o que o autor chama de infovigilância, e os métodos que analisam o que está disponível sem envolver as pessoas ele chama de passivos.

O exemplo fundador é de saúde, e não de desinformação. Um estudo de 2006, do próprio Eysenbach, mostrou correlação entre buscas sobre gripe no Google e casos de gripe na semana seguinte, no Canadá, e que as buscas precediam em uma semana as consultas a médicos-sentinela (Eysenbach, 2009). Como o Google não compartilhava dados de busca com pesquisadores externos, o estudo comprou um anúncio acionado por palavra-chave, que devolvia estatísticas de busca e de clique. O detalhe importa por duas razões. Mostra que as medidas de difusão dependem de dados que as plataformas controlam, e Lazer et al. (2018) dizem que a cooperação das plataformas com os pesquisado-

res será crítica para entender a escala do problema e a eficácia das intervenções. Mostra ainda o que esses dados são: registros do que as pessoas fazem ao mensurar buscar e clicar, e não do que dizem, o que o autor apresenta como vantagem sobre inquéritos, por serem diretos e livres de viés de recordação e de desejabilidade social (Eysenbach, 2009).

A vantagem vem com um custo que o próprio autor registra. A busca isolada é ambígua quanto à intenção de quem busca (alguém que procura “resfriado” não necessariamente tem os sintomas), e as contagens de resultados dos buscadores não são exatas nem confiáveis (Eysenbach, 2009). Na linguagem do Capítulo 3, a forma da resposta, a palavra buscada, não diz a função. O autor admite ainda a seta no outro sentido: mudanças nos padrões de informação podem afetar a saúde da população, como num “surto” de desinformação ou numa campanha, e em 2009 escreve que não existiam metodologias sólidas para identificar e acompanhar automaticamente a prevalência da informação que leva a menores taxas de vacinação (Eysenbach, 2009). A frase é de 2009, e não se afirma que continue verdadeira; marca o ponto de partida do campo.

A palavra infodemia ficou conhecida em 2020. Em 15 de fevereiro daquele ano, o diretor-geral da Organização Mundial da Saúde (OMS) levantou a preocupação de que a epidemia vinha acompanhada de uma infodemia (Calleja et al., 2021), e a organização a define como um excesso de informação de qualidade variável, em ambientes digitais e físicos, durante um evento agudo de saúde, que leva a confusão, a comportamentos de risco e à perda de confiança nas autoridades. Eysenbach (2020) escreve que cunhou infodemiologia em 2002 e, ao longo de vinte anos, infodemia e infovigilância; o artigo de 2009, que foi uma das bases consultadas, usa os termos infodemiologia e infovigilância, e não traz a palavra infodemia.

Zielinski (2021) reduz as características primárias de uma infodemia a duas, o volume de informação gerado e a velocidade com que aparece, e trata as formas de informação errada como características secundárias, ainda que as mais importantes. O autor pergunta se um grande volume de informação em grande parte verdadeira e útil, surgindo depressa, seria uma infodemia, e responde que, para quem gerencia informação, qualquer surto de informação de grande escala e fluxo rápido precisa ser gerenciado. O quadro do artigo associa ao volume problemas de localizar, armazenar, avaliar a qualidade, dar visibilidade e validar, e à velocidade, problemas de avaliar, filtrar, aplicar, rastrear a história e evitar desperdício, como pesquisa repetida sem necessidade. Os exemplos vêm sobretudo da literatura científica: o aumento de mais de 400% nas postagens de preprints no medRxiv (de 586 nas últimas 15 semanas de 2019 para 2.572 nas primeiras 15 de 2020), que o artigo cita de outro trabalho.

Isso reorganiza a pergunta. Os Capítulos 1 a 7 trataram de enunciados falsos e do que os sustenta. A infodemiologia mede um fenômeno de outra ordem: o volume e a velocidade da informação, verdadeira ou falsa, e o que isso faz com a capacidade de quem recebe de separar uma da outra. O falso é parte do problema, e não sua definição. Eysenbach (2020) acrescenta uma cautela que o capítulo adota. No início de uma pandemia, os “fatos” são mais bem chamados de “BETs” (best evidence at the time), a melhor evidência no momento; as recomendações baseadas neles estão sujeitas a mudança, e combater a infodemia espalhando fatos é mais fácil de dizer que de fazer quando não está claro quais são os fatos. O autor dá como exemplos as recomendações sobre máscaras, o uso de certos medicamentos e as regras de distanciamento. A consequência para qualquer medida de “informação falsa” em saúde é aqui proposta: o rótulo de falso depende de quem verifica, com que evidência e em que momento, e o Capítulo 2 tratou dessa dependência para as agências de checagem.

11.2 Medidas de difusão: alcance, exposição, reação e a estrutura da cascata

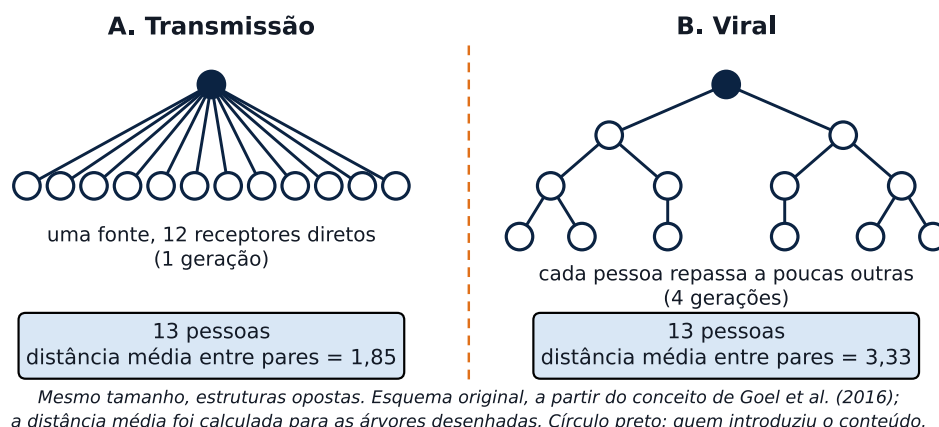
Medir a difusão exige decidir o que conta como uma pessoa que “recebeu” e o que conta como uma que “passou adiante”. Os estudos do capítulo usam a unidade da cascata: o conjunto de repasses ligados, por uma cadeia ininterrupta, a uma origem comum. Uma cascata pode ter tamanho um, se ninguém repassou a emissão original, e um rumor pode ter várias cascatas, tantas quantas as vezes em que foi introduzido de modo independente (Vosoughi et al., 2018). Cada cascata é descrita por quatro números: o tamanho, que é o número de usuários envolvidos; a profundidade, que é o número de saltos de retuíte a partir da origem, em que um salto é um retuíte de um usuário novo; a amplitude máxima, que é o maior número de usuários envolvidos em qualquer profundidade; e a viralidade estrutural.

A viralidade estrutural é utilizada por Goel et al. (2016) e responde a um problema de interpretação. Quando se diz que algo “viralizou”, entende-se que cresceu de pessoa a pessoa, como um vírus; mas um conteúdo pode chegar a milhões por uma única grande transmissão, como a de uma conta com dezenas de milhões de seguidores, sem que ninguém o repasse a outro. O tamanho final não distingue os dois casos. A profundidade também falha, porque uma única corrente longa, numa transmissão enorme, dá profundidade alta. Os autores definem a viralidade estrutural como a distância média entre todos os pares de pessoas da árvore da cascata, medida que a química matemática conhece como índice de Wiener, e que não depende de suposição sobre o proces-

so que gerou a difusão. Numa transmissão pura, ela fica perto de 2, o mínimo; quanto mais gerações, mais alta. A Figura 11.1 mostra duas árvores de 13 pessoas e a diferença que o número registra.

FIGURA 11.1

Transmissão e difusão viral: duas árvores de mesmo tamanho



Nota. Esquema original, a partir do conceito de Goel et al. (2016); sem dados. A distância média entre todos os pares foi calculada para as duas árvores desenhadas; numa transmissão com muitas pessoas, o valor tende a 2.

O conjunto de Goel et al. reúne cerca de 1 bilhão de eventos de difusão no Twitter, de julho de 2011 a junho de 2012: tuítes com links de sites populares de notícias, vídeos, imagens e petições. O que acharam desfaz a imagem do viral. O tamanho médio das árvores é 1,3 e, cerca de 99% das adoções são das raízes ou de seus seguidores imediatos; só 0,025% das árvores chegam a 100 nós ou mais, 219.855 em cerca de 1 bilhão. Entre essas há quase todas as estruturas, da transmissão pura às ramificadas, mas a viralidade estrutural mediana é baixa e pouco varia com o tamanho: menos de 3 para imagens e vídeos, mesmo com 10.000 repostagens ou mais; entre 7 e 8 para petições; e abaixo de 8 para as maiores cascatas de notícias, de 3.000 repostagens. A correlação entre tamanho e viralidade estrutural é 0,2 para notícias, 0,04 para petições e essencialmente zero para imagens e vídeos, e os autores concluem que, se a popularidade se relaciona de modo consistente com alguma característica, é com o tamanho da maior transmissão. Sobre as causas dizem que só podem especular: suspeitam da falta de grandes canais de transmissão para petições e, para imagens e vídeos, de celebridades com dezenas de milhões de seguidores.

O resultado vale com o limite que eles mesmos marcam: restringe-se à difusão de informação no Twitter, e sem reconstruir a sequência dos eventos e sem uma medida de viralidade a mera popularidade permite concluir pouco sobre a importância relativa dos mecanismos viral e de transmissão. Também a re-

construção é uma inferência: cada adotante é ligado ao amigo, isto é, à conta que ele segue, que adotou o conteúdo mais recentemente antes dele, e os autores dizem inferir corretamente o pai em cerca de 95% dos casos.

Três termos organizam o que as medidas contam, e aqui eles são fixados porque os Capítulos 8 e 10 os usam e a literatura os mistura.

DEFINIÇÃO PROPOSTA

Alcance, exposição e reação.

Alcance é o conjunto de pessoas a quem uma emissão é entregue ou que podem vê-la; depende da rede de contatos e da regra de entrega da plataforma (Capítulo 8). *Exposição* é o que uma pessoa efetivamente viu. *Reação* é o que se faz depois de ver: curtir, comentar, repassar ou ignorar. Skinner não usa os termos; a separação é proposta aqui e serve para dizer, de cada medida, qual dos três ela conta.

Os estudos do capítulo contam sobretudo reação, e quando falam de exposição a inferem. Goel et al. presumem a exposição pela lista de contas seguidas; Mosleh e Rand (2022) constroem o escore de exposição de cada usuário a elites que mentem a partir de quem ele segue, e lembram que exposição e compartilhamento são construtos fundamentalmente diferentes: a maior parte das pessoas compartilha só uma pequena fração do que vê. O alcance, por sua vez, é o que a plataforma fixa, e o Capítulo 8 viu que o ranqueamento altera o alcance de cada emissão de modo desigual (Huszár et al., 2022). A Tabela 11.1 diz, de cada estudo do capítulo, o que conta como difusão e o que fica de fora.

Nenhuma dessas medidas diz o que a emissão fez ao comportamento de quem a viu. Lazer et al. (2018) escrevem que exposição não equivale a impacto e que a avaliação do efeito da exposição a notícias falsas sobre o comportamento político a médio e longo prazo é essencialmente inexistente na literatura. O capítulo volta a esse ponto na seção 11.6.

TABELA 11.1*O que cada estudo conta como difusão e o que fica de fora*

Estudo	O que conta	Unidade	Mede	Não Mede
Eysenbach (2009)	Buscas (demanda) e publicações (oferta) com uma palavra-chave ou conceito	Termo ou conceito em um conjunto de páginas ou buscas	Ocorrências e buscas	Quem viu; o que fez com a informação; a intenção da busca
Goel et al. (2016)	Repostagens de links, ligadas à conta seguida que repostou antes	Árvore de cada conteúdo	Reação (repostar); a exposição é presumida	O que a plataforma entregou; quem viu sem repostar; a veracidade do conteúdo
Vosoughi et al. (2018)	Retuítas de uma afirmação checada	Cascata de rumor	Reação (retuitar), em número de usuários	Quem viu sem retuitar; quais rumores ninguém checou
Cinelli et al. (2020)	Pessoas que publicam sobre o tema; posts e reações por classe de veículo	Autores por dia; posts e reações por link	Reação (publicar, comentar, reagir)	Exposição; a veracidade de cada afirmação (o rótulo é do veículo)
Massarani et al. (2021)	Engajamento total do link (curtidas, comentários, compartilhamentos, reações)	Link entre os cem de maior engajamento	Reação agregada, sem distinguir o tipo	Alcance; links fora dos cem; quem reagiu

Nota. Elaboração própria, a partir dos métodos dos artigos. A coluna “Mede” usa os três termos da definição acima.

11.3 O falso se espalha mais rápido que o verdadeiro?

A resposta mais citada da área a essa pergunta é a de Vosoughi et al. (2018), e vale conhecer o desenho antes do resultado. Os autores reuniram as cascatas de rumores que seis organizações de checagem haviam verificado e que circularam no Twitter de 2006 a 2017: cerca de 126.000 histórias, tuitadas por cerca de 3 milhões de pessoas mais de 4,5 milhões de vezes, classificadas como verdadeiras ou falsas com 95 a 98% de concordância entre as organizações. A coleta partiu de respostas, em inglês, a tuítes que traziam o link de uma das páginas de checagem; o tuíte original e seus retuítas formam a cascata. Os autores evitam a expressão “notícia falsa”, por considerarem ter se tornado políti-

camente polarizada e sem relação com a veracidade. Alternativamente, usam “verdadeiro” e “falso” e dizem que não afirmam nada sobre a intenção de quem divulga a informação.

O falso se difundiu significativamente mais longe, mais rápido, mais fundo e mais amplamente que o verdadeiro em todas as categorias de informação, e o efeito foi mais pronunciado para o falso político que para o falso sobre terrorismo, desastres naturais, ciência, lendas urbanas ou finanças. O verdadeiro raramente chegou a mais de 1.000 pessoas, e o 1% superior das cascatas falsas chegava rotineiramente a entre 1.000 e 100.000; ademais, o verdadeiro levou cerca de seis vezes mais tempo que o falso para chegar a 1.500 pessoas. Num modelo da probabilidade de retuitar, o falso foi 70% mais provável de ser retuitado, controlando idade da conta, atividade e número de seguidores e seguidos de quem o originou, e se a conta era verificada. Um dado pede cuidado: uma fração maior de rumores falsos teve entre 1 e 1.000 cascatas, e uma fração maior de rumores verdadeiros teve mais de 1.000. O verdadeiro, portanto, teve com mais frequência muitas origens independentes, e o resultado do falso é um resultado de alcance por cascata.

Os autores propõem uma explicação e a apresentam com cautela. Em todas as métricas de novidade, calculadas contra os tuítes a que os usuários haviam sido expostos nos 60 dias anteriores, o falso era mais novo que o verdadeiro, embora os usuários possam não o ter percebido como novo. Nas respostas, segundo um léxico de emoções, o falso suscitou mais surpresa e nojo, e o verdadeiro, mais tristeza, antecipação, alegria e confiança. Mas eles escrevem que não podem afirmar que a novidade cause retuítes nem que seja a única razão pela qual o falso é mais retuitado, e que a explicação comportamental exige pesquisa com contato direto com os usuários. Sobre os robôs, os autores encontraram dados de que eles aceleraram a difusão de notícias verdadeiras e falsas na mesma taxa, e concluem que o falso se espalha mais porque humanos, e não robôs, têm mais probabilidade de espalhá-lo.

Os limites são do desenho. Entram só rumores que alguém checou, e as organizações podem escolher certos tipos de rumor ou chamar atenção para eles; os autores testaram a objeção com 13.240 cascatas que estudantes, trabalhando de modo independente, investigaram e rotularam como verdadeiras, falsas ou mistas, com 90% de concordância e kappa de Fleiss de 0,88 (índice de concordância entre avaliadores que desconta a concordância esperada por acaso), e acharam resultados quase idênticos. Tudo é Twitter, em inglês, até 2017. E “mais longe” quer dizer mais usuários novos que retuitaram, e não mais pessoas convencidas. Há ainda uma tensão aparente com o estudo de Goel et al. da seção anterior. Vosoughi et al. dizem que o falso não se espalhou por transmissão simples, e sim por difusão de pessoa a pessoa, num processo de

ramificação viral; Goel et al. mostram que, em média, as grandes cascatas são dominadas por transmissão. As duas coisas cabem juntas, porque uma compara falso e verdadeiro e a outra descreve o conjunto, e lembram que “viral” é medida de estrutura, que se compara, e não rótulo que se aplica a tudo o que é grande.

Além de Vosoughi et al., dois outros estudos do capítulo fazem a mesma pergunta com outra unidade, e a Tabela 11.2 os põe lado a lado.

TABELA 11.2

Três estudos, três unidades: o que cada um comparou entre o confiável e o falso

Estudo	Unidade comparada	Quem classificou	Plataforma e período	O que se achou
Vosoughi et al. (2018)	Afirmação checada (cascata de rumor)	Seis organizações de checagem	Twitter, 2006-2017	O falso chegou mais longe, mais rápido, mais fundo e mais amplamente; o efeito foi maior para o político
Cinelli et al. (2020)	Link, rotulado pelo veículo (confiável ou questionável)	Media Bias/Fact Check	Cinco plataformas, 1.º de janeiro a 14 de fevereiro de 2020, sobre Covid-19	Posts e reações de fontes confiáveis e questionáveis crescem no mesmo ritmo em cada plataforma; a amplificação relativa varia com a plataforma
Massarani et al. (2021)	Link de vacina, entre os cem de maior engajamento	Os próprios autores, com categorias adaptadas de outro trabalho	Facebook, Twitter, Pinterest e Reddit, em português, 2018-2019 e 2020	Proporção de desinformação igual nos dois períodos; engajamento médio maior para a desinformação só em 2020

Nota. Os resultados vêm dos artigos; as unidades e os critérios, nas seções 11.2 e 11.5. Não são medidas comparáveis entre si.

Nenhum trabalho que reexamine, com outro método, a comparação de Vosoughi et al. foi lido. Três trabalhos previstos no plano deste capítulo não puderam ser obtidos por via legal e automatizada (a lista está no Apêndice D). O capítulo, portanto, não diz o que concluem, e o resultado de Vosoughi et al. aparece aqui sem o exame que trabalhos posteriores possam ter feito dele.

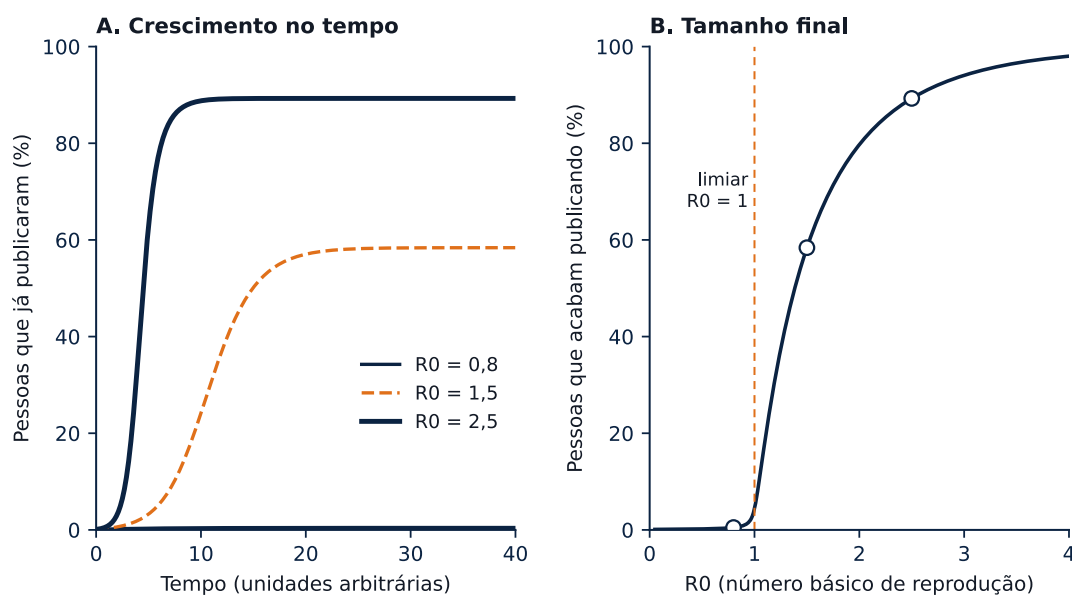
■ DADO	Afirmações checadas e falsas chegaram mais longe e mais rápido que as verdadeiras no Twitter, de 2006 a 2017 (Vosoughi et al., 2018). Em cinco plataformas e nos primeiros 45 dias de 2020, links de veículos rotulados como questionáveis e confiáveis cresceram no mesmo ritmo dentro de cada plataforma, e a amplificação relativa variou entre elas (Cinelli et al., 2020). Em 94 e 89 links de vacina, a proporção de desinformação foi praticamente a mesma em 2018–2019 e 2020, e o engajamento médio da desinformação superou o do verificado só em 2020 (Massarani et al., 2021).
■ INTERPRETAÇÃO	“O falso se espalha mais” é afirmação sobre uma unidade (a afirmação checada), uma plataforma e um período. A mesma pergunta com outra unidade, o veículo, ou em outra plataforma, não dá o mesmo resultado, e nenhum dos estudos contraria o outro, porque medem coisas diferentes.
■ HIPÓTESE	Quando o falso se espalha mais, é porque propriedades do conteúdo, como a novidade, controlam respostas de repassar que a audiência já emitia, no sentido dos Capítulos 8 e 9. É compatível com os dados de novidade, surpresa e nojo de Vosoughi et al., que não têm teste causal; os próprios autores dizem que não podem afirmar que a novidade cause retuítes.

11.4 Modelos epidemiológicos aplicados à informação

Os modelos epidemiológicos dividem a população em compartimentos, suscetíveis, infectados e recuperados (SIR); a pessoa infectada contagia cada contato suscetível com certa probabilidade; e o número básico de reprodução, R_0 , é o número médio de casos secundários que um infectado produz. Na epidemiologia, $R_0 = 1$ é o limiar: abaixo dele a doença se extingue, acima se espalha (Cinelli et al., 2020). Goel et al. (2016) escrevem R_0 como o produto do número médio de contatos de cada nó pela probabilidade de contágio por contato. A Figura 11.2 mostra o que o limiar faz, numa simulação própria, com parâmetros ilustrativos e sem dados.

FIGURA 11.2

O limiar de R_0 num modelo SIR simples



Nota. Simulação original: modelo SIR de mistura homogênea (todos têm o mesmo contato com todos), com 0,1% de pessoas iniciais e recuperação de 1 por unidade de tempo. Parâmetros ilustrativos; nenhum dado de plataforma, tema ou período. Painel B: tamanho final calculado para cada R_0 .

Cinelli et al. (2020) aplicaram o modelo a 1.342.103 posts e 7.465.721 comentários de 3.734.815 usuários em cinco plataformas, de 1.º de janeiro a 14 de fevereiro de 2020. O que cresce é o número de autores que publicam sobre o tema, e o “infectado” é quem passa a publicar depois de exposto a ele; na abordagem dos autores, R_0 maior que 1 indicaria a emergência de uma infodemia. Ajustaram dois modelos: um fenomenológico (EXP), que busca reproduzir os dados sem supor um mecanismo de crescimento específico, e o SIR (*Susceptible, Infected, Recovered*) padrão. Todos os valores estimados ficam acima de 1 (são “supercríticos”), mesmo considerando os intervalos de confiança: com o EXP, de 1,42 a 1,52 no Gab¹ e de 2,02 a 2,64 no Instagram; com o SIR, de 2,2 a 2,5 no Gab e de 4,0 a 5,1 no Twitter, e cerca de 100 no Instagram, valor que os autores dizem estar muito além do observado em qualquer epidemia real e que atribuem a um salto abrupto de novos autores que modelos contínuos não explicam.

¹ Rede americana de microblog.

Dois cuidados valem para esses números. O primeiro é o que R_0 é aqui: um parâmetro de crescimento do número de pessoas que publicam sobre um assunto numa plataforma, em 45 dias, e não uma propriedade de uma afirmação fal-

sa nem de uma pessoa. O segundo é que o ajuste do modelo à curva não estabelece o mecanismo. Goel et al. (2016) escrevem que replicar alguns resultados empíricos com um modelo teórico não implica que ele represente o processo gerador verdadeiro, embora seja possível descartar alguns modelos.

O uso mais instrutivo do modelo, na literatura consultada, é o de Goel et al., que o empregaram para descartar hipóteses. O modelo em rede aleatória é descartado porque produziria tamanhos concentrados em torno da média, e não a cauda pesada; com rede livre de escala e baixa infecciosidade (r em torno de 0,5, isto é, subcrítico), o modelo reproduz a probabilidade de uma emissão ser popular, a viralidade estrutural média e a correlação com o tamanho, mas não a diversidade de estruturas observada. Os autores concluem que a noção de ponto de inflexão (*tipping point*) é, em grande parte, irrelevante para os eventos virais que estudam, o que vai contra a ideia comum de que a informação “viraliza” ao passar um limiar: tudo o que observaram, inclusive os maiores eventos, cabe num modelo de baixa infecciosidade em que, numa rede com poucos polos de muitos seguidores, um conteúdo encontra com facilidade um nó que o sustenta.

Em linguagem comportamental, o que o modelo chama de infecção é uma resposta verbal: publicar ou repassar uma mensagem na presença da emissão de outra pessoa. O modelo trata essa resposta como evento que a exposição desencadeia com probabilidade igual para todos os suscetíveis e não tem termo para a consequência que a mantém. É o que os Capítulos 6, 8 e 9 abriram: a difusão é reforçada pelos efeitos sobre as reações de uma audiência específica, e a probabilidade de difundir varia conforme audiência (Capítulo 6), o conteúdo (Capítulo 8) e história de quem repassa (Capítulo 9).

CONCEITO EM USO

Macrocontingência (Glenn, 2004), retomada do Capítulo 9

O Capítulo 9 (seção 9.7) distinguiu a metacontingência da macrocontingência. Macrocontingência é a relação entre uma prática cultural e a soma agregada das consequências do macrocomportamento que a constitui; esses efeitos cumulativos, escreve Glenn (2004), não podem controlar, como consequências operantes, o comportamento de cada indivíduo. Uma cascata de 10.000 repasses é um efeito agregado desse tipo: nenhum dos 10.000 foi mantido pelo tamanho da cascata. O que mantém cada repasse é o retorno que a pessoa recebe de sua audiência. Para Glenn, a engenharia cultural eficaz exige identificar as macrocontingências que produzem efeitos indesejáveis e alterar as contingências operantes ou as metacontingências relevantes. Leitura aqui proposta: reduzir o tamanho das cascatas pede mudar o que mantém o repasse de cada um, e não o tamanho.

Isso não torna a escala irrelevante. O que preocupa a saúde pública é o efeito agregado, e é nele que o modelo epidemiológico mede. A distinção é de nível: o fenômeno se mede no nível da população e se explica, na leitura aqui proposta, no entrelaçamento das contingências de diferentes pessoas.

■ DADO	Modelos SIR se ajustam à curva de crescimento do número de autores sobre um tema em cinco plataformas (Cinelli et al., 2020), e um SIR em rede livre de escala com baixa infecciosidade reproduz a distribuição de tamanhos das cascatas no Twitter, mas não a diversidade de suas estruturas (Goel et al., 2016).
■ INTERPRETAÇÃO	O modelo descreve o crescimento sem descrever o que mantém cada resposta de repassar; seu parâmetro de contágio reúne, num número, alcance e probabilidade de reação.
■ HIPÓTESE	A probabilidade de repasse é produto das contingências do Capítulo 9 (reação da audiência, indignação, pertencimento ao grupo) e varia entre pessoas e conteúdos de um modo que o modelo de mistura homogênea não capta. É compatível com o fracasso do modelo em reproduzir a diversidade estrutural, mas esse fracasso tem outras explicações, e nenhuma fonte lida testa a hipótese.

11.5 A infodemia de Covid-19, no mundo e no Brasil

Dois estudos sobre a infodemia de Covid-19, um em escala mundial e um sobre o Brasil, possibilitam avançar na compreensão referente a este processo. O de Cinelli et al. (2020) compara a Covid-19 em cinco ambientes. O Twitter, o Instagram e o YouTube são plataformas convencionais; o Gab, que os autores descrevem como financiado por doações, com muito pouco controle de conteúdo e com base de usuários considerada de extrema direita, e o Reddit, agregador de notícias com filtragem coletiva. Os autores classificaram os links como confiáveis ou questionáveis segundo o Media Bias/Fact Check, que rotula veículos e não artigos, e explicitam esse limite. Em cada plataforma, os posts e as reações de fontes questionáveis e confiáveis crescem no mesmo ritmo, e o fator de escala varia fortemente com a plataforma. No Gab, os posts de fontes questionáveis eram cerca de 70% dos de fontes confiáveis, e as reações, cerca de 3 vezes maiores. O coeficiente de amplificação relativa, que é a razão entre as reações médias por post de fonte questionável e de fonte confiável, é cerca de 1 no Twitter, 0,35 no YouTube, 0,55 no Reddit e 3,9 no Gab. Os autores concluem que a informação de fontes confiáveis ou questionáveis não difere de modo significativo nos padrões de difusão, que o Gab é o ambiente mais

suscetível à desinformação. Ainda, o que impulsiona a difusão são peculiaridades de cada plataforma e a dinâmica de grupo de quem se engaja com o tema, mostrando quão forte é o efeito ancorado na história de quem repassa.

Os limites são os da coleta. O estudo cobre os primeiros 45 dias, e o primeiro dado do Twitter é de 27 de janeiro, por limites da interface de programação da plataforma; a coleta por fluxo do Twitter é cerca de 1% do volume, e a do Instagram foi feita à mão. A amplificação é relativa a cada plataforma e depende do rótulo do veículo, não da veracidade de cada afirmação. E os autores notam, na introdução, que estudos haviam apontado que a informação falsa pode se espalhar mais rápido e mais amplamente, mas que isso poderia ser efeito específico da plataforma, o que seus dados não contrariam.

Para o Brasil, Massarani et al. (2021) usaram o BuzzSumo, ferramenta de monitoramento que mostra que conteúdos sobre um tema geraram mais engajamento nas redes, para coletar os cem links (o máximo que a ferramenta permite) com a palavra “vacina” que tiveram maior engajamento em 2020. Destes, 94 links válidos, em português - depois de excluídos cinco links de veículos de Portugal e um link quebrado - foram comparados com os de 22 de maio de 2018 a 21 de maio de 2019, que geraram 89 links, com os mesmos filtros. Os autores aplicaram a noção de engajamento a partir da soma de compartilhamentos, comentários, curtidas e reações no Facebook, de compartilhamentos no Twitter, de compartilhamentos e comentários no Reddit e de compartilhamentos no Pinterest, mas ressaltaram que a definição se restringiu aos protocolos das redes. Classificaram a acurácia de cada link, comparando as informações citadas com artigos científicos, fontes oficiais e outros veículos, em verificado, não verificável (como opiniões pessoais, enquetes e conteúdo de ficção) ou desinformação, esta subdividida em tipos como conteúdo enganoso, fabricado e conexão falsa, com categorias adaptadas de proposta de outros autores. Classificaram também o emissor de cada link: profissional, se tinha política editorial definida, instituição responsável, autores e equipe identificados e citação de fontes, ou não profissional. A Tabela 11.3 reúne os números.

A média de engajamento por link aumentou 8,6 vezes, de 41.405 para 354.321. A proporção de desinformação praticamente não mudou, nem a de informação verificada. O que mudou foi o perfil. Em 2018–2019, a desinformação era sobretudo conteúdo fabricado, isto é, 100% falso e criado para enganar ou prejudicar (66,7%), emitido por veículos não profissionais (83,3%); em 2020, a maior parte era conexão falsa (84,6%), e os veículos profissionais passaram a ser os principais emissores (61,5%). A média de engajamento da desinformação, que em 2018–2019 era menor que a do verificado (36.200 contra 40.483), foi maior em 2020 (412.223 contra 338.019).

A leitura que os autores fazem é que o maior engajamento da desinformação sugere maior potencial de viralização na pandemia, e ligam a conexão falsa a manchetes sensacionalistas e à dependência de receita publicitária baseada em engajamento, apoiados em outros trabalhos. Ela é registrada como interpretação deles: o artigo não mede receita nem a edição das manchetes. A leitura converge com a hipótese do Capítulo 8, de que o retorno por atenção seleciona formas de manchete, e este estudo não a testa. Com os percentuais, 2018–2019 corresponde a cerca de 12 itens de desinformação e 2020 a cerca de 13, uma conta aqui reconstituída, mas o artigo apoia-se em medidas descritivas. Diferenças de médias entre 12 e 13 itens, com a variação de engajamento que as redes têm, não sustentam muito mais que a descrição apresentada.

TABELA 11.3

Links de vacina de maior engajamento nas redes, 2018–2019 e 2020

Medida	2018–2019	2020
Links válidos	89	94
Engajamento total	3.685.100	33.306.200
Média de engajamento por link	41.405	354.321
Informação verificada	80,9%	81,9%
Desinformação	13,5%	13,8%
Não verificável	5,6%	4,3%
Média de engajamento: verificada	40.483	338.019
Média de engajamento: desinformação	36.200	412.223
Tipo de desinformação: fabricado / conexão falsa	66,7% / 8,3%	15,4% / 84,6%
Principal emissor da desinformação	Veículos não profissionais (83,3%)	Veículos profissionais (61,5%)

Nota. Valores do artigo. Em 2018–2019 havia também 25% de conteúdo enganoso entre a desinformação; em 2020, nenhum. “Conexão falsa”: títulos, manchetes e elementos visuais que contradizem o conteúdo (Quadro 1 do artigo).

Os autores veem, nos dois períodos, instrumentalização política do debate sobre vacinas, com exemplos de campos diferentes. Em 2018–2019 descrevem conteúdo falso e checagens sobre a vacina contra meningite e um conteúdo falso que atribuía a um político estrangeiro uma fala que ele não fez; em 2020, a Coronavac foi a vacina mais citada em conteúdo desinformativo (23,1%), embora não liderasse as menções gerais, e os autores leem as dúvidas sobre ela como apoiadas em questões políticas, e não em evidência científica. A leitura é dos autores: o artigo mede quais vacinas foram citadas, e não por quem as cita o faz. Um retorno ao Capítulo 1 (seção 1.6) mostra que ele reúne os demais dados brasileiros sobre desinformação.

■ DADO	Nos primeiros 45 dias de 2020, o ritmo de crescimento de posts e reações de fontes questionáveis e confiáveis foi o mesmo dentro de cada plataforma, e a amplificação relativa variou de cerca de 0,35 a 3,9 (Cinelli et al., 2020). Nos links de vacina de maior engajamento no Brasil, o engajamento médio aumentou 8,6 vezes entre 2018–2019 e 2020, a proporção de desinformação ficou praticamente igual, e seu perfil mudou, de conteúdo fabricado por veículos não profissionais para manchete que contradiz o conteúdo, emitida sobretudo por veículos profissionais (Massarani et al., 2021).
■ INTERPRETAÇÃO	Nos dois estudos, a infodemia de Covid-19 aparece como aumento de volume e de engajamento e mudança de perfil, e não como aumento da proporção de falso. A plataforma e o período mudam o resultado.
■ HIPÓTESE	A mudança de perfil, de fabricado por não profissionais para manchete enganosa por profissionais, reflete a contingência de retorno por atenção a que os veículos profissionais estão sujeitos (Capítulo 8). É a interpretação dos autores do estudo, sem teste; seria preciso medir a edição das manchetes antes e depois.

11.6 O que a infodemiologia propõe fazer

Eysenbach (2020) propõe quatro pilares para o manejo da infodemia: monitorar a informação (a infovigilância), construir letramento em saúde e em ciência, estimular o refinamento do conhecimento por processos como a checagem e a revisão por pares, e traduzir o conhecimento de modo oportuno e sem distorções políticas ou comerciais. Ele descreve a crise em quatro níveis: ciência, política e prática, mídia de notícias e mídias sociais, em camadas de tamanho proporcional à quantidade de informação que cada grupo de atores gera, e diz que o modelo é opinião de especialista, e não o quadro da OMS.

A consulta da OMS que o artigo de Calleja et al. (2021) relata, reuniu 110 participantes de mais de 35 países e produziu, por julgamento estruturado de especialistas, uma agenda de pesquisa com cinco eixos, 166 perguntas sugeridas e 49 de alta prioridade. Os eixos são medir e monitorar o impacto; detectar sinais e entender a difusão e o risco; responder com intervenções; avaliar as intervenções e fortalecer a resiliência; e promover o desenvolvimento e a aplicação de intervenções e ferramentas. Os autores dizem que os efeitos nocivos da infodemia não podem ser geridos só pelas abordagens correntes de comunicação, engajamento comunitário e mensagens, e que a resposta exige inter-

venção em vários níveis: indivíduo, comunidade, meio, plataforma e política. Escrevem que campanhas de comunicação podem aumentar a visibilidade de mensagens, mas muitas vezes não são eficazes para desmentir afirmações falsas, e que metade das organizações de checagem não trabalha com profissionais de saúde. Listam, entre os desafios do ambiente informacional, a amplificação computacional de mensagens polarizadoras, a microsegmentação e o enfraquecimento do jornalismo local.

Zielinski (2021) acrescenta duas posições. A primeira é que importa manter a distinção entre *misinformation* e *disinformation*, porque as soluções provavelmente diferem: grandes doses de boa informação ajudam contra a primeira, e a segunda usa os mesmos canais da boa informação. Distingue ainda a ‘malinformação’, informação baseada na realidade usada para infligir dano, definição do Conselho da Europa que o artigo reproduz. A segunda é que as ações corretivas precisam de salvaguardas de direitos humanos, para não levarem à censura e à restrição da liberdade de expressão, e que a tecnologia sozinha não resolve.

O que o capítulo leva para a Parte IV são três pontos. O primeiro é a distinção entre as coisas que uma intervenção pode alterar, e a definição de alcance, exposição e reação organiza a pergunta: a regra de entrega da plataforma (Capítulo 12, seção 12.6) altera o alcance; sinais e treino alteram a reação (seções 12.2 a 12.5); em conjunto, agir sobre as consequências do alcance e da exposição da desinformação podem reduzir danos e, a reação pode retroalimentar a própria entrega. O segundo é que a pauta do campo é de perguntas, e não de respostas: a agenda da OMS é fruto de julgamento de especialistas, e aqui ela é lida como mapa do que se ignora; e o que se ignora pode envolver condições que direcionaria respostas com menor grau de desinformação. O terceiro é que nenhum dos estudos de difusão lidos mede o que a informação fez ao comportamento de saúde das pessoas, como vacinar-se ou aderir a uma medida. Mede a difusão, e a difusão é o desfecho que o campo mede com mais facilidade. Isso limita o que se pode dizer sobre a eficácia de qualquer intervenção avaliada só por menor difusão, e o Capítulo 12 retoma a limitação ao avaliar as intervenções.

11.7 Perguntas para ler um estudo de difusão

Para quem usa o capítulo na leitura de estudos, de notícias sobre estudos ou de relatórios de plataformas, as perguntas que as seções anteriores sugerem são as seguintes.

1. Qual é a unidade que se compara: afirmação checada, link, veículo, tópico, conta?

2. Quem classificou o verdadeiro e o falso, e com que critério (verificador, rótulo do veículo, os próprios autores)?
3. O que se conta: alcance, exposição ou reação? O que foi inferido, e a partir de quê?
4. A medida é de nível (quanto) ou de estrutura (como, por quantas gerações)? O tamanho final diz pouco sobre a estrutura.
5. Qual a base de comparação, a plataforma, o idioma e o período?
6. O estudo mede o que a informação fez ao comportamento de quem a viu, ou só o quanto se espalhou?

Síntese

Infodemiologia é, na proposta de Eysenbach, a ciência da distribuição e dos determinantes da informação em meio eletrônico, com o fim de informar a saúde pública; infodemia, no uso da OMS, é o excesso de informação de qualidade variável durante um evento agudo de saúde. Para Zielinski, a infodemia se define por volume e velocidade, e o falso é característica secundária. O caso da Lombardia, no relato dos autores que o contam, mostra que informação não falsa pode ter efeito danoso por seu momento e pelo que as pessoas fazem com ela (11.1).

As medidas de difusão contam sobretudo reação (repostar, retuitar, engajar), inferem a exposição pela lista de quem se segue e deixam o alcance à regra da plataforma. A viralidade estrutural, a distância média entre os pares de uma cascata, mostra que, no Twitter de 2011–2012, quase tudo é pequeno e quase tudo é transmissão, e que o tamanho de uma cascata diz pouco sobre sua estrutura (11.2). O falso foi mais longe e mais rápido que o verdadeiro no Twitter de 2006 a 2017, entre afirmações checadas, e outros dois estudos, com outras unidades, não dão o mesmo resultado; as revisões que o plano previa não foram lidas (11.3).

Os modelos epidemiológicos descrevem o crescimento e permitem descartar hipóteses, como a do ponto de inflexão, mas tratam a resposta de repassar como evento desencadeado pela exposição, sem termo para a consequência que a mantém; a escala é do nível da população, e a explicação, na leitura aqui proposta, é de contingências entre pessoas (11.4). Na Covid-19, nos primeiros 45 dias de 2020, o ritmo de crescimento foi o mesmo para fontes confiáveis e questionáveis dentro de cada plataforma, e nos links de vacina no Brasil o engajamento aumentou 8,6 vezes, com a proporção de desinformação praticamente igual e o perfil mudado (11.5). O campo propõe monitorar, letrar, refinar e traduzir, em vários níveis, e sua agenda é de perguntas; nenhum estudo de difusão lido mede o efeito sobre o que as pessoas fazem (11.6).

A distinção entre alcance, exposição e reação, que os Capítulos 8 e 10 anunciaram, fica fixada na seção 11.2 e serve para ler as medidas dos Capítulos 8 e 9, que contam um dos três sem separá-los, e para a ressalva do Capítulo 10 de que a janela efetiva de uma publicação depende de quem a vê. O Capítulo 12 parte do que aqui ficou aberto: o que se pode mudar, em quatro níveis, e com que evidência.

Referências

- Calleja, N., AbdAllah, A., Abad, N., . . . Purnat, T. D. (2021). A public health research agenda for managing infodemics: Methods and results of the first WHO infodemiology conference. *JMIR Infodemiology*, 1(1), Article e30979. <https://doi.org/10.2196/30979>
- Cinelli, M., Quattrocioni, W., Galeazzi, A., Valensise, C. M., Brugnoti, E., Schmidt, A. L., Zola, P., Zollo, F., & Scala, A. (2020). The COVID-19 social media infodemic. *Scientific Reports*, 10(1), Article 16598. <https://doi.org/10.1038/s41598-020-73510-5>
- Eysenbach, G. (2009). Infodemiology and infoveillance: Framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. *Journal of Medical Internet Research*, 11(1), Article e11. <https://doi.org/10.2196/jmir.1157>
- Eysenbach, G. (2020). How to fight an infodemic: The four pillars of infodemic management. *Journal of Medical Internet Research*, 22(6), Article e21820. <https://doi.org/10.2196/21820>
- Glenn, S. S. (2004). Individual behavior, culture, and social change. *The Behavior Analyst*, 27(2), 133–151. <https://doi.org/10.1007/bf03393175>
- Goel, S., Anderson, A., Hofman, J., & Watts, D. J. (2016). The structural virality of online diffusion. *Management Science*, 62(1), 180–196. <https://doi.org/10.1287/mnsc.2015.2158>
- Huszár, F., Ktena, S. I., O'Brien, C., Belli, L., Schlaikjer, A., & Hardt, M. (2022). Algorithmic amplification of politics on Twitter. *Proceedings of the National Academy of Sciences*, 119(1), Article e2025334119. <https://doi.org/10.1073/pnas.2025334119>
- Lazer, D. M. J., Baum, M. A., Benkler, Y., ... & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094–1096. <https://doi.org/10.1126/science.aao2998>
- Massarani, L. M., Leal, T., Waltz, I., & Medeiros, A. (2021). Infodemia, desinformação e vacinas: a circulação de conteúdos em redes sociais antes e depois da COVID-19. *Liinc em Revista*, 17(1), Article e5689. <https://doi.org/10.18617/liinc.v17i1.5689>
- Mosleh, M., & Rand, D. G. (2022). Measuring exposure to misinformation from political elites on Twitter. *Nature Communications*, 13(1), Article 7144. <https://doi.org/10.1038/s41467-022-34769-6>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>
- Zielinski, C. (2021). Infodemics and infodemiology: A short history, a long future. *Revista Panamericana de Salud Pública*, 45, 1. <https://doi.org/10.26633/rpsp.2021.40>



PARTE IV

Intervenções

- 12** Intervenções: o que se pode mudar, em quatro níveis, e com que evidência

12

Intervenções: o que se pode mudar, em quatro níveis, e com que evidência

Um estudo de campo com anúncios no YouTube serve de ponto de partida. Um vídeo de inoculação, visto por 30 segundos ou mais, foi seguido, em até 24 horas, de uma única pergunta: qual técnica de manipulação usava uma postagem fictícia. A proporção de acertos foi maior entre quem viu o anúncio do que num grupo de comparação, e os autores descrevem o resultado como um aumento médio de cerca de 5% no reconhecimento. O tamanho do efeito, medido pelo h de Cohen (uma medida da diferença entre duas proporções, em que 0,2 é, por convenção, um efeito pequeno), foi de 0,09, abaixo do menor efeito de interesse que os próprios autores tinham pré-registrado (0,10) (Roozenbeek et al., 2022).

É um dos estudos de intervenção mais cuidadosos entre os lidos: seis experimentos randomizados, nos cinco primeiros com hipóteses pré-registradas, e um estudo de campo com mais de 22 mil respostas. É também, por isso, um bom ponto de partida. O que ele mede é o reconhecimento de uma técnica por quem viu um anúncio. Não mede se a pessoa compartilha menos, nem o que sua audiência faz com o que ela compartilha, nem o que acontece uma semana depois; os autores dizem que o estudo de campo avaliou a suscetibilidade, e não o comportamento. Entre o que o estudo mede e o que os Capítulos 3 a 11 trataram como aquilo que mantém a fala falsa, ou classificada por instância externa como discurso de ódio, há uma distância, e o capítulo existe para medi-la, intervenção por intervenção.

A análise feita nos capítulos anteriores pôs a manutenção da fala em contingências, e não na informação: a audiência que reage (Capítulo 9), a regra de entrega que decide o alcance (Capítulo 8), a comunidade verbal que reforça e

protege o falante (Capítulo 6) e as práticas culturais que reúnem tudo isso (Capítulos 9 e 10). Cada uma sugere um ponto de intervenção, e o capítulo as organiza em quatro níveis, que a agenda da OMS lista de modo próximo (indivíduo, comunidade, meio, plataforma e política; Calleja et al., 2021): o indivíduo, que é o que a pessoa lê, vê ou aprende; a interação e o grupo, que é o que se faz diante de quem fala; a plataforma, que é o que a regra de entrega e a moderação mudam; e a cultura, que são as práticas e as metacontingências.

A Figura 12.1 põe lado a lado onze estudos de intervenção lidos na íntegra e diz, de cada um, o que foi medido. A contagem é própria, feita sobre a figura. Oito estudos mediram julgamento ou desempenho em tarefa; cinco, comportamento observado numa plataforma ou fórum; quatro, comportamento observado com seguimento de uma semana ou mais; um só, o de Guess et al. (2020), mediu as quatro coisas; e nenhum avaliou uma intervenção em nível cultural. A distribuição não é acaso. O indivíduo é o nível em que é barato testar, porque se mostra um material a alguém e se pergunta o que achou. Os outros níveis exigem acesso a contas, a fóruns e ao ranqueamento, e é neles que a medida passa a ser o que as pessoas fazem.

Três cautelas orientam a leitura. A primeira é que nenhuma intervenção foi testada aqui. Os estudos lidos são de terceiros, interpretados à luz da análise operante, e quando se diz que uma intervenção “ensina a discriminar” ou “age sobre a contingência”, a frase é leitura aqui proposta e vem marcada como hipótese. Também não se leram as revisões e réplicas posteriores sobre correção de afirmações, nem o estudo do Twitter que um dos artigos citados descreve (seção 12.3); onde isso importa, o texto diz. A segunda é a medida. O Capítulo 1 mostrou que a resposta a uma pesquisa é comportamento sob controle da situação em que foi dada, e o Capítulo 11 que nenhum estudo de difusão lido mede o que a informação fez ao comportamento das pessoas. O mesmo vale aqui: julgar uma manchete numa tela não é compartilhá-la, e compartilhá-la não é mudar o que se faz; a correspondência entre dizer e fazer, que o Capítulo 7 (seção 7.4) tratou como efeito de contingência, e não como dado, vale para a intenção declarada de compartilhar. A terceira é o critério. O Capítulo 10 usou o mesmo para o que alarga e para o que estreita a janela de Overton, e o capítulo o usa para as intervenções: a mesma pergunta, o mesmo tipo de medida e a mesma ressalva, qualquer que seja o campo político de quem emite o conteúdo que se quer reduzir.

O percurso segue os níveis e, dentro do indivíduo, o momento da intervenção. A seção 12.1 trata de corrigir depois, com a correção e o aviso; a 12.2, de corrigir antes, com a inoculação lida como treino discriminativo; a 12.3, dos estímulos que sinalizam precisão. A 12.4 trata do que se faz diante de quem fala (contra-fala, compromisso público e a hipótese dos autoclíticos de fonte); a

12.5, da redução de preconceito por equivalência e por histórias em quadrinhos; a 12.6, das contingências da plataforma. A 12.7 reúne o que se entende aqui por planejamento cultural e os limites éticos de quem planeja.

FIGURA 12.1

O que cada estudo de intervenção mediu

	Julgamento ou desempenho em tarefa	Intenção declarada de compartilhar	Comportamento observado (plataforma ou fórum)	Seguimento de uma semana ou mais
Indivíduo: o que a pessoa lê, vê ou aprende				
Nyhan e Reifler (2010) <i>correção dentro de notícia simulada</i>	●	–	–	–
Freeze et al. (2021) <i>aviso retrospectivo de desinformação</i>	●	–	–	–
Roozenbeek e van der Linden (2019) <i>jogo de inoculação</i>	●	–	–	–
Roozenbeek et al. (2022) <i>vídeos de inoculação</i>	●	●	–	–
Guess et al. (2020) <i>dicas para identificar notícias falsas</i>	●	●	●	●
Dixon e Lemke (2007) <i>treino de equivalência (adultos)</i>	●	–	–	–
Mizael et al. (2021b) <i>treino de equivalência (crianças)</i>	●	–	–	●
Interação e grupo: o que se faz diante de quem fala				
Hangartner et al. (2021) <i>contra-fala por empatia, no Twitter</i>	–	–	●	●
Tsipursky et al. (2018) <i>compromisso público com a verdade</i>	–	–	●	●
Plataforma: o que a regra de entrega e a moderação mudam				
Álvarez-Benjumea e Winter (2018) <i>remoção de comentários hostis (fórum simulado)</i>	–	–	●	–
Nyhan et al. (2023) <i>rebaixar fontes de mesma inclinação (Facebook)</i>	●	–	●	●
Cultura: práticas e metacontingências				
Nenhuma intervenção em nível cultural foi avaliada nos estudos lidos para este livro.				

● a variável foi medida (o resultado pode ter sido nulo). – não foi medida.

Mapa original, a partir dos estudos lidos (Capítulos 4 e 12). Seguimento: Guess et al., cerca de 20 dias (EUA); Hangartner et al. e Tsipursky et al., quatro semanas; Nyhan et al., três meses; Mizael et al., seis semanas. Medida no mesmo dia ou em até 24 horas não conta como seguimento.

Nota. Mapa original, a partir dos estudos lidos (Capítulos 4 e 12); não traz dados novos. A marca registra que a variável foi medida, e o resultado pode ter sido nulo. Seguimento: cerca de 20 dias em Guess et al. (nos Estados Unidos), quatro semanas em Hangartner et al. e em Tsipursky et al., três meses em Nyhan et al. (2023) e seis semanas em Mizael et al. (2021b); medida no mesmo dia ou em até 24 horas não conta como seguimento. As leituras de cada célula estão nas seções 12.1 a 12.6.

12.1 Corrigir depois: a correção, o aviso e o “efeito rebote”

A intervenção mais comum diante de uma afirmação falsa é a correção: depois que a afirmação circulou, alguém a desmente. Nyhan e Reifler (2010) a testaram em quatro experimentos online, com estudantes de graduação de uma universidade católica do Meio-Oeste dos Estados Unidos, no outono de 2005 e na primavera de 2006. Os participantes liam uma notícia simulada que

trazia uma afirmação enganosa de um político, com ou sem correção, e diziam depois se concordavam com ela. Os autores escolheram um delineamento entre sujeitos para maximizar o efeito das correções, porque medir antes e depois ancoraria as respostas na primeira. Definem percepções errôneas como crenças falsas ou não comprovadas sobre o mundo, e chamam de backfire o caso em que a pessoa passa a apoiar a opinião original ainda mais fortemente.

Os resultados não são uniformes. No Estudo 1 (130 participantes), a correção do relatório Duelfer sobre armas de destruição em massa no Iraque não reduziu a percepção errônea no conjunto; entre os conservadores ($n = 33$), a concordância de que o Iraque tinha tais armas subiu de 32% no controle para 64% com a correção, e entre os não conservadores ($n = 97$) foi de 22% para 13%. No experimento sobre cortes de impostos, a concordância dos conservadores de que os cortes aumentaram a receita foi de 36% para 67% com a correção ($n = 60$), e a dos não conservadores, de 31% para 28% ($n = 136$). Em outros dois experimentos, o resultado foi outro: na segunda versão do experimento sobre armas de destruição em massa, a correção tornou os conservadores mais propensos a acreditar que o Iraque não tinha essas armas, e os autores dizem que não está claro por quê; no experimento sobre células-tronco, a correção funcionou para conservadores e moderados, e não para liberais, sem backfire. Atribuir a notícia ao *New York Times* ou ao FoxNews.com não alterou o efeito da correção em nenhum dos três experimentos do Estudo 2.

Os autores colocam as ressalvas no próprio artigo. Todos os resultados de backfire vieram de conservadores, e sem mais estudos é impossível saber se liberais e conservadores reagem de modo diferente a correções; seria valioso replicar com amostra não universitária. Uma análise post hoc, que os autores descrevem como levar os dados ao limite, sugere que a correção falhou entre os conservadores para quem o Iraque era o problema mais importante, mas só 34 respondentes (oito à direita do centro) o apontaram. E escrevem que as percepções errôneas contemporâneas sobre política são um alvo móvel, que pode mudar depressa, como sugere a diferença entre os dois experimentos sobre armas de destruição em massa.

A correção também chega como aviso. Freeze et al. (2021) estudaram o que chamam de efeito da “verdade contaminada”: o aviso de desinformação aplicado a notícias autênticas, que reduz a credibilidade delas e prejudica a memória política. Num experimento online com 434 participantes da MTurk, todos viram um vídeo de quatro minutos da C-SPAN, com discursos de deputados, responderam a 22 perguntas sem relação com ele e foram sorteados para seis condições. Liam um artigo noticioso fabricado sobre o vídeo, ora vago (controle), ora com os fatos corretos, ora com parte dos fatos alterados, e recebiam ou não o aviso “alguma informação do artigo que você leu era imprecisa”.

sa”. Os autores chamam o aviso de válido quando aponta corretamente a desinformação, e de inválido quando recai sobre um artigo sem fatos alterados. Depois todos fizeram um teste de reconhecimento de 20 fatos do vídeo e avaliaram a credibilidade do vídeo e do artigo. O aviso fez o artigo parecer 0,59 ponto menos crível (numa escala de 5), em todas as condições, inclusive quando o artigo era exato. O efeito sobre a memória só foi significativo em um dos três testes.

<p>■ DADO</p>	<p>Em quatro experimentos com estudantes, a correção embutida na notícia fez a concordância dos conservadores com a afirmação enganosa subir em dois (armas de destruição em massa, Estudo 1; cortes de impostos) e funcionou entre eles nos outros dois (armas, segunda versão; células-tronco, para conservadores e moderados e não para liberais). Entre os não conservadores, a concordância caiu nos dois temas em que os números são relatados (Nyhan & Reifler, 2010).</p>
<p>■ INTERPRETAÇÃO</p>	<p>O estudo mede a concordância com uma afirmação, numa pesquisa, logo depois da leitura. Não diz o que a pessoa faz com a afirmação nem por quanto tempo a resposta se mantém. Os resultados de backfire vêm só de percepções errôneas mantidas por conservadores, nos temas escolhidos, e os autores dizem não poder afirmar que os campos políticos diferem; o estudo não sustenta nem “conservadores rejeitam correções” nem “correções saem pela culatra”.</p>
<p>■ HIPÓTESE</p>	<p>Os autores leem o backfire como contra-argumentação diante de informação incongruente com a preferência (interpretação deles, sem teste). Acrescentam-se aqui duas leituras rivais, igualmente sem teste neste desenho. A primeira é a do Capítulo 6: a correção é um estímulo aversivo para quem tem a afirmação ligada à sua comunidade, e discordar mais é resposta de contracontrole, ou esquiva dele. A segunda é a da repetição (Capítulo 3, seção 3.5): a correção reapresenta a afirmação, e instruções do tipo “não acredite que todos os membros do grupo X são Y” podem aumentar a saliência do grupo e acrescentar atributos à categoria (Dixon et al., 2003). Separar as três leituras pediria variar se a pessoa declarou antes a posição em público e se a correção repete ou não o enunciado.</p>

O aviso retrospectivo válido, dado depois da leitura, ajudou pouco. O escore médio de memória foi de 59% no controle, 46% com desinformação e sem aviso e 52% com desinformação e aviso válido, e o escore de desinformação passou de 33% para 25%; a diferença entre 46% e 52% não foi significativa a 0,05 ($p = 0,06$). Os autores concluem que seus achados sugerem evitar avisos gerais,

porque o uso indiscriminado pode reduzir a credibilidade de fontes válidas e levar as pessoas a descartar informação útil. Notam também que o aviso partiu do pesquisador, sem fonte claramente especificada, e que seria interessante ver se o raciocínio motivado altera o efeito quando ele vem de outra fonte.

Lido aqui, o resultado de Freeze et al. tem a forma de uma generalização de estímulo. Um aviso sem alvo (“isto pode ser desinformação”) não diz qual relação do artigo está em jogo, e o leitor responde reduzindo a confiança no conjunto, inclusive no que é válido. Um aviso específico, que localizasse o enunciado e a razão, dá ocasião a outra resposta. É hipótese, e o estudo não a testa: todos os avisos foram gerais, e nenhum variou a especificidade. A medida, mais uma vez, é de julgamento: credibilidade e memória, logo depois da leitura.

O que a seção deixa como dado é pouco, e é útil por isso. Corrigir depois, no desenho dos estudos lidos, tem efeito incerto, que variou com o tema e não variou com a fonte atribuída à notícia nos experimentos em que isso foi manipulado, e não foi medido além da sessão. A hipótese de que a correção seja aver-siva para quem tem a afirmação ligada à comunidade (Capítulo 6) é compatível com os resultados, e nenhum estudo a testou. A seção seguinte trata do que se fez para evitar, em vez de corrigir.

12.2 Corrigir antes: inoculação como treino discriminativo

A inoculação inverte a ordem da correção. Em vez de desmentir uma afirmação depois de ela ter circulado, expõe-se a pessoa, antes, a uma versão enfraquecida do ataque. Roozenbeek et al. (2022) definem a inoculação psicológica por dois componentes: um aviso prévio que induz a ameaça percebida de um ataque iminente às atitudes, e a exposição a uma microdose enfraquecida de desinformação que traz refutação preventiva. A teoria é atribuída pelos autores a McGuire, nos anos 1960, e segue uma analogia com a imunização médica; não é conceito da análise do comportamento, e a leitura que o capítulo faz dela é aqui proposta.

O primeiro estudo lido é um jogo. Em *Bad News*, o jogador assume o papel de produtor de notícias falsas e aprende seis técnicas de produção de desinformação: polarização, apelo a emoções, teorias da conspiração, trolagem, desvio de culpa (a desacredibilização de opositores, que desvia a atenção de acusações de viés atacando a fonte da crítica) e personificação de contas. O jogo é gratuito, roda no navegador e leva cerca de 15 minutos; o jogador ganha seguidores e credibilidade ao usar as estratégias e perde credibilidade ou seguidores ao escolher opções alinhadas ao comportamento jornalístico ético (Roozenbeek & van der Linden, 2019). A amostra era de conveniência, recrutada por comu-

nicado de imprensa da universidade, com 14.266 pares pré-pós completos coletados em três meses de 2018; era autosselecionada e enviesada para homens (75%) e para pessoas de algo a muito liberais (59%).

A medida era a confiabilidade, de 1 a 7, de tuítes e manchetes avaliada antes e depois do jogo, em seis itens, dois deles controles de notícias reais. A confiabilidade atribuída caiu em personificação (de 3,00 para 2,30; $d = 0,36$), conspiração (de 2,47 para 1,97; $d = 0,35$) e descridibilização (de 2,37 para 1,92; $d = 0,30$), e os itens de controle praticamente não mudaram, o que os autores leem como indício de que o jogo não torna os participantes apenas mais céticos. O item de polarização, coletado post hoc por falha técnica de armazenamento (885 pares), teve efeito menor ($d = 0,16$). O efeito médio sobre os itens falsos foi de 2,61 para 2,06 ($d = 0,52$) e foi maior entre quem os julgava mais confiáveis no pré-teste.

Os autores marcam as limitações. Por razão ética não randomizaram quem jogaria: não há grupo de controle tradicional, e dizem que explicações alternativas não podem ser totalmente descartadas; estudos futuros devem modelar o decaimento do efeito, para saber se e quando seriam necessárias doses de reforço. Há uma ressalva a acrescentar: o efeito maior entre quem julgava os itens mais confiáveis no pré-teste é o padrão esperado também da regressão à média, e sem grupo de controle não há como separá-la.

O estudo de 2022 é de outro porte. Cinco vídeos curtos tratam de linguagem emocionalmente manipuladora, incoerência, falsas dicotomias, bode expiatório e ataques *ad hominem*; cada um traz aviso prévio, refutação preventiva da técnica e microdose em exemplos inócuos, humorísticos, apolíticos e fictícios. Nos estudos 1 a 6, os participantes foram designados ao acaso a assistir a um vídeo de inoculação de 1,5 minuto ou a um vídeo neutro, e depois avaliaram dez postagens sintéticas, manipulativas ou neutras, em quatro medidas, entre elas a disposição a compartilhar. O resultado principal é a discriminação (*discernment*): a diferença entre as médias dos estímulos manipulativos e dos neutros. Os estímulos manipulativos nem sempre eram falsos; eram feitos para usar claramente uma técnica de manipulação.

Nos estudos 1 a 5 ($n = 5.416$, amostras de cotas nacionais dos Estados Unidos), as hipóteses pré-registradas foram sustentadas em 16 de 20 medidas. O d de Cohen para o reconhecimento da técnica foi de 0,28 (bode expiatório) a 0,68 (falsas dicotomias) e, para o compartilhamento, na forma de discriminação, de 0,10 (incoerência, não significativo) a 0,22 (falsas dicotomias). Um ano depois, o estudo 6 ($n = 1.068$) replicou o vídeo de linguagem emocional, com $d =$

0,34 para compartilhamento. Para o efeito menor nos vídeos de bode expiatório e incoerência, a explicação dos autores é que o grupo de controle já distinguia melhor essas técnicas, isto é, a linha de base era maior.

No estudo de campo, dois vídeos rodaram como anúncios no YouTube, que os espectadores podiam pular; cerca de 967 mil usuários o assistiram por 30 segundos ou mais, e 30% dos escolhidos ao acaso receberam, em até 24 horas, uma única pergunta de teste dentro do próprio YouTube, em que deviam identificar a técnica usada numa postagem fictícia (quatro alternativas, só uma correta). O grupo de comparação, pareado em demografia, não viu o vídeo e respondeu a uma das perguntas; foram coletadas 22.632 respostas. A proporção de acertos foi maior no tratamento ($h = 0,09$, abaixo do menor efeito de interesse pré-registrado, $h = 0,10$). Os autores relatam custo médio de cerca de US\$ 0,05 por visualização e dizem que os vídeos foram mostrados a cerca de 5,4 milhões de pessoas em 15 dias.

As limitações declaradas são as que importam para a leitura. O efeito foi medido em até 24 horas (mediana de 18,4 horas), a duração não foi estudada, os vídeos e as amostras eram dos Estados Unidos, e o estudo de campo avaliou a suscetibilidade, e não o comportamento. Cada participante do campo viu um único item de teste, e os pares de estímulos manipulativo e neutro não são espelhos exatos, podendo diferir em outras dimensões. Os autores resumem, citando um estudo que não foi lido, que induzir atenção à acurácia levou usuários do Twitter a compartilhar mais fontes de alta qualidade, e dizem que como reduzir a suscetibilidade em ambientes reais segue uma questão aberta. Não encontraram interações consistentes entre a condição e as covariáveis demográficas e ideológicas.

A leitura aqui proposta é que a inoculação, no estudo de 2022, tem a forma de um treino discriminativo com exemplares da técnica, e não do conteúdo. A propriedade comum aos exemplos é a técnica (a linguagem emocional, a falsa dicotomia); o conteúdo varia e, no teste, é outro. O que se mede é a diferença entre a avaliação de uma postagem que usa a técnica e a de uma postagem neutra, de mesmo conteúdo, e é isso o que o conceito de discriminação pede. A Figura 12.2 põe o procedimento nessa forma, com o que o estudo mede em traço contínuo e o que se acrescenta aqui em tracejado.

CONCEITO EM USO

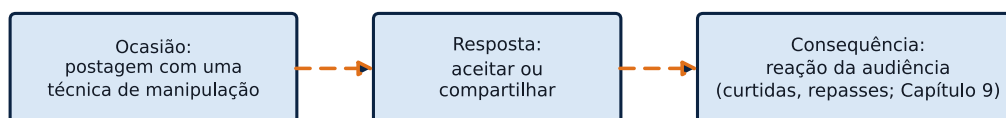
Treino discriminativo (Skinner, 1957; Shea et al., 2023)

Em todo comportamento verbal sob controle de estímulos, o estímulo prévio estabelece a ocasião em que a resposta tende a ser reforçada e, por um processo de discriminação operante, passa a ser a ocasião em que a resposta tende a ser emitida (Skinner, 1957). No tato, o controle se estabelece reforçando a resposta o mais consistentemente possível na presença de um estímulo, com muitos reforçadores diferentes ou com um reforçador generalizado, e o controle resultante é pelo estímulo. O *treino de múltiplos exemplares*, na descrição de Shea et al. (2023), apresenta um conjunto de estímulos que amostra a faixa dos estímulos-alvo e das topografias de resposta, e continua até que a resposta fique sob controle das características dos exemplos e também de estímulos novos que as compartilham. Leitura aqui proposta: o treino discriminativo é o procedimento que aproxima uma resposta de uma propriedade do estímulo, e não do estímulo particular; o teste de que funcionou é a resposta a itens novos que compartilham a propriedade, com outro conteúdo.

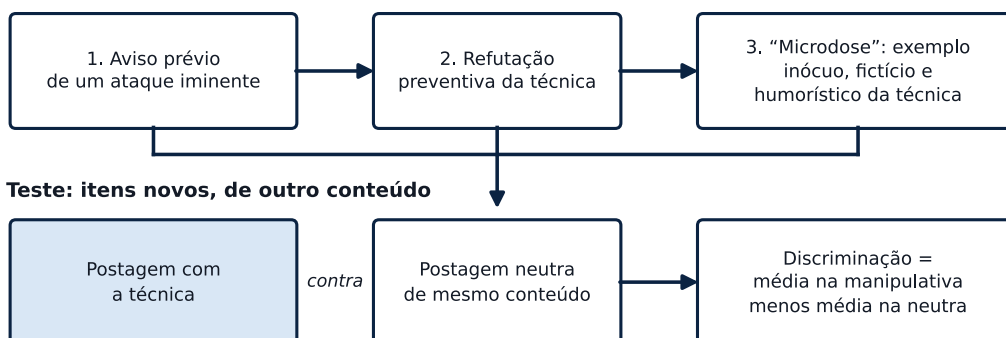
FIGURA 12.2

Inoculação por vídeo lida como treino discriminativo

Fora do treino (leitura aqui proposta, sem teste direto)



Inoculação por vídeo (Roizenbeek et al., 2022): o procedimento



Esquema original, sem dados. Traço contínuo: o procedimento do estudo. Tracejado: leitura aqui proposta, que a medida do estudo não alcança (o estudo de campo não mediu o compartilhar; a consequência social não entra no treino).

Nota. Esquema original, sem dados. Traço contínuo: o procedimento do estudo. Tracejado: leitura aqui proposta, que a medida do estudo não alcança; o estudo de campo não mediu o compartilhar, e a consequência social não entra no treino.

A leitura tem dois pontos fracos, que convém dizer. O primeiro é que, nos vídeos, não há consequência diferencial programada para discriminar: o espectador vê, ouve e responde depois. O que o procedimento tem é instrução e exemplo, e o termo “treino” é, em boa parte, analogia. Em *Bad News* há consequência, em seguidores e credibilidade, mas é por usar a técnica, e não por reconhecer-lá; o teste mede o outro repertório, o de avaliar, e a transferência de um ao outro é suposição. O segundo é que a faixa tracejada da Figura 12.2 é a que importa para a análise: a técnica aparece, fora do estudo, numa ocasião em que há audiência, e a resposta de compartilhar tem a consequência que o Capítulo 9 descreveu. Nada no treino mexe nessa consequência.

■ DADO	No jogo <i>Bad News</i> , sem grupo de controle, a confiabilidade atribuída a itens com três das técnicas caiu com d de 0,30 a 0,36, e a dos itens reais não mudou (Roozenbeek & van der Linden, 2019). Nos vídeos, seis experimentos randomizados sustentaram 16 de 20 hipóteses pré-registradas nos estudos 1 a 5, e o estudo de campo teve $h = 0,09$, abaixo do menor efeito de interesse, medido em até 24 horas, nos Estados Unidos e como suscetibilidade (Roozenbeek et al., 2022).
■ INTERPRETAÇÃO	Uma exposição breve a exemplos de uma técnica melhora, logo depois, a avaliação de itens novos que a usam, sem tornar a pessoa apenas mais cética. O efeito é pequeno no campo, e o que se mede é julgamento, e não comportamento.
■ HIPÓTESE	A inoculação funciona como treino com múltiplos exemplares da técnica, e a resposta de avaliar generaliza a conteúdos novos (leitura aqui proposta). Uma leitura rival é a de que o espectador recebeu uma regra, e a responde enquanto a regra está saliente. As duas preveem o efeito logo depois; diferem na duração e no que acontece quando a regra compete com a reação da audiência. Separá-las pediria medir, com seguimento, o compartilhar numa plataforma em que a mesma postagem recebe ou não reação.

12.3 Estímulos que sinalizam precisão

Uma terceira família de intervenções, além da correção (seção 12.1) e da inoculação (12.2), não ensina antes nem corrige depois: põe, no momento em que a pessoa lê, um estímulo que sinaliza o que olhar. A que foi lida em detalhe é a lista de dicas que o Facebook promoveu no topo do feed em 14 países, em abril de 2017, desenvolvida com a organização First Draft: “Dicas para identificar notícias falsas”, com 10 estratégias na versão dos Estados Unidos e 6 na da Índia.

dia. Guess et al. (2020) a descrevem como regras de decisão simples que ajudam a distinguir notícias principais de falsas, o que chamam de discriminação, sem exigir que o leitor pesquise ou reflita a fundo sobre cada notícia.

Nos Estados Unidos, o estudo foi um painel online de duas ondas, conduzido pela YouGov depois das eleições legislativas de 2018 (onda 1, $N = 4.907$; onda 2, $N = 4.283$), com participantes distribuídos ao acaso para ler ou não as dicas. A medida principal foi a acurácia percebida (de 1 a 4) de manchetes principais, hiperpartidárias e falsas, que circularam em fontes reais ou no Facebook ou WhatsApp nos seis meses anteriores, parte delas classificada como falsa por ao menos uma agência de checagem. Na primeira onda, as dicas reduziram a acurácia percebida das notícias falsas em quase 0,2 ponto na escala de 4 pontos ($ITT = -0,196$, o efeito sobre todos os designados) e em quase 0,3 entre os efetivamente tratados ($ATT = -0,299$); cerca de 66% passaram nas perguntas de compreensão.

A intervenção teve efeito negativo pequeno sobre a crença em notícias principais ($ITT = -0,046$), mas aumentou a discriminação entre principais e falsas ($ITT = 0,146$), uma melhora de 26,5% em relação ao grupo de controle. Com mediana de 20 dias entre as ondas, o efeito sobre a acurácia percebida de notícias falsas seguiu distinguível de zero, mas atenuou-se em mais da metade ($ITT = -0,080$). Na medida binária, a proporção que avaliava manchetes falsas como muito ou algo precisas caiu de 32% no controle para 24% na primeira onda e, na segunda, de 33% para 29%.

Na amostra online da Índia, a intervenção reduziu a acurácia percebida das notícias falsas ($ITT = -0,126$) e melhorou a discriminação em 17,3%; só 28% passaram nas perguntas de compreensão, contra 66% nos Estados Unidos. Na amostra presencial, aleatória do cadastro eleitoral de quatro distritos de Uttar Pradesh e com aplicação oral, os autores não encontraram evidência de efeito; 11% dessa amostra usavam WhatsApp, contra 90% da amostra online. Na segunda onda, não encontraram evidência confiável de efeito em nenhuma das duas amostras da Índia.

O que o estudo não encontrou importa tanto quanto o que encontrou. Os autores dizem que não há evidência suficiente de que a intervenção tenha mudado a intenção declarada de compartilhar notícias falsas nem medidas comportamentais de consumo de notícias; as intenções de compartilhar notícias principais aumentaram, e as de hiperpartidárias diminuíram. Entre as ressalvas: os efeitos foram modestos e decaíram; houve efeito negativo pequeno sobre a acurácia percebida das notícias principais; muitos usuários reais do Facebook provavelmente ignoraram o link; e a ausência de mudança no consumo real talvez se deva a que os hábitos de informação são arraigados. Com dados de

monitoramento de navegação, estimam, num cálculo que descrevem como hipotético, que os falsos positivos cairiam de 6,1% para 4,9% das histórias consumidas.

Dois comentários dos autores situam a intervenção. Comparam o efeito às respostas a rótulos de “disputado” do Facebook em estudo anterior (ITT = -0,236) e a um aviso geral (ITT = -0,079); esses estudos não foram lidos. E notam que tentar identificar notícias falsas pelo escrutínio da manchete difere da abordagem dos checadores profissionais, que usam a leitura lateral de outras fontes, e que estudos futuros devem esclarecer se o efeito vem de maior preocupação com a acurácia ou de heurísticas mais eficazes.

Há dois outros resultados de sinalização de acurácia nas fontes lidas. Num dos estudos de Ceylan et al. (2023), julgar a acurácia de uma manchete antes de decidir compartilhá-la reduziu o compartilhamento de manchetes falsas, mas não aumentou a discriminação dos usuários mais habituais mais do que a dos menos habituais; o estudo foi feito com 839 participantes, e a tarefa era um botão de compartilhar simulado. E Freeze et al. (2021), na seção 12.1, mostram o preço do sinal sem alvo.

<p>■ DADO</p>	<p>Nos Estados Unidos, a lista de dicas reduziu em cerca de 0,2 ponto (de 4) a acurácia percebida de notícias falsas, aumentou em 26,5% a discriminação, e o efeito caiu em mais da metade com mediana de 20 dias; na Índia, o efeito foi menor na amostra online e não foi encontrado na presencial nem na segunda onda. Não houve evidência suficiente de mudança na intenção declarada de compartilhar notícias falsas nem no consumo observado (Guess et al., 2020).</p>
<p>■ INTERPRETAÇÃO</p>	<p>A dica muda o julgamento de uma manchete, por pouco tempo e onde foi entendida (66% nos Estados Unidos, 28% na amostra online da Índia) e onde o meio de circulação é o do estudo (11% de WhatsApp na amostra presencial, 90% na online). Não muda, nos dados, o que as pessoas fazem com as notícias.</p>
<p>■ HIPÓTESE</p>	<p>A dica é um estímulo verbal que descreve uma contingência: uma regra, no vocabulário do Capítulo 3 (seção 3.7). Regras controlam enquanto estão salientes e a resposta é barata; sem consequência programada por aplicá-las, e com a reação da audiência (Capítulo 9) mantendo a resposta de compartilhar, o decaimento que os autores encontraram é o esperado, e o hábito, no sentido de Ceylan et al., deixa pouco espaço para elas. É leitura aqui proposta; o mecanismo, dizem os próprios autores, está por esclarecer.</p>

12.4 Diante de quem fala: contra-fala, compromisso público e autoclíticos de fonte

Entre quem lê e a plataforma que entrega há a interação: o que alguém faz, na mesma rede, diante de quem emitiu um enunciado falso ou classificado, por instância externa, como discurso de ódio. Dois estudos lidos avaliam intervenções desse tipo. A seção termina com uma hipótese aqui proposta que nenhum estudo testa.

Hangartner et al. (2021) fizeram um experimento de campo no Twitter. Selecionaram, por dicionário de termos xenófobos, análise de sentimento e classificação manual, usuários que tinham enviado tuítes xenófobos ou racistas em inglês, excluindo contas verificadas, retuítes, tuítes possivelmente sarcásticos e contas provavelmente de bots ou menores. Cada usuário foi sorteado para uma de três contra-falas de aplicação única (empatia, aviso de consequências, humor) ou para o controle (40%), sem intervenção; a contra-fala vinha, em até 24 horas, de um perfil neutro controlado por humano, criado cerca de quatro semanas antes, com histórico apolítico e cerca de 100 seguidores. A de empatia buscava despertar empatia (dizer, por exemplo, que é doloroso ver esse tipo de linguagem); a de aviso lembrava que amigos e familiares podem ver o tuíte; a de humor usava um meme de animal com legenda pedindo para parar de tuitar. O campo foi de 23 de novembro de 2020 a 17 de janeiro de 2021, e a amostra analítica, depois da perda de 230 contas privadas ou suspensas, foi de 1.350 usuários (humor 284, aviso 269, empatia 259, controle 538).

A análise foi por intenção de tratar, isto é, cada usuário conta no grupo a que foi sorteado, tenha ou não recebido a mensagem (5,5% dos tuítes foram apagados por quem os enviou entre o sorteio e a aplicação), com regressões separadas para cada um dos seis desfechos. Como são muitos desfechos e tratamentos, os autores dão também valores de P ajustados (pelo método de Benjamini-Hochberg, que corrige o excesso de resultados significativos ao acaso que testes múltiplos produzem). A empatia reduziu o número de tuítes xenófobos nas quatro semanas seguintes em 0,10 desvio-padrão ($P = 0,056$; P ajustado = 0,112) e o total de tuítes em 0,13 ($P = 0,039$; ajustado = 0,112), e aumentou a exclusão do tuíte original em 0,21 ($P = 0,009$; ajustado = 0,053), cerca de 8,4 pontos percentuais. Em contagem, os usuários do grupo empatia enviaram, em média, 1,3 tuíte xenófobo a menos e 91,6 tuítes a menos no total. O teste F conjunto, que examina de uma vez todos os desfechos, rejeitou a hipótese nula de ausência de efeito para a empatia ($P = 0,017$), mas não para o humor ($P = 0,285$) nem para o aviso de consequências ($P = 0,633$).

Os autores dizem que, ao contrário das hipóteses pré-registradas, não encontraram efeitos significativos da empatia nos outros três desfechos, e que a análise separada dos dois componentes da parcela de tuítes xenófobos é exploratória; o desfecho pré-registrado, a própria parcela, não mudou. Concluem que os resultados moderam as expectativas sobre as contra-falas mais comuns usadas por organizações e recomendam estudar intervenções repetidas e atentar para efeitos de arrefecimento (*chilling*) “para além do discurso de ódio”, expressão deles. O estudo foi aprovado por comitê de ética e pré-registrado. A premissa, que os autores apoiam em outros trabalhos, é que a moderação de conteúdo pode suprimir fala valiosa e ilícita e dispersar, em vez de reduzir, o que chamam de discurso de ódio, e que a contra-fala busca reduzir o ódio por persuasão de quem o emite, e não suprimir a expressão.

O segundo estudo é de outra natureza. Tsipursky et al. (2018) descrevem o *Pro-Truth Pledge*, proposta em que signatários, cidadãos e figuras públicas, se comprometem com 12 comportamentos correlacionados com a orientação à verdade, agrupados em compartilhar, honrar e encorajar a verdade; entre eles, verificar a informação antes de aceitá-la e compartilhá-la, retratar-se se não puder verificá-la e pedir a aliados que retratem informação refutada. Na análise dos autores, o compromisso funciona como estímulo discriminativo verbal, uma regra que descreve uma contingência em que comportamentos verazes teriam consequências positivas e os não verazes, negativas. Para figuras públicas, o desenho prevê a reputação como consequência: defensores voluntários monitoram, um comitê emite aviso à imprensa em caso de desacato e a figura é listada no site como em desacato; o artigo descreve o mecanismo como desenho e não relata uso nem sanções efetivas.

O compromisso foi lançado em dezembro de 2016 e, em maio de 2018, tinha mais de 7.500 signatários, dos quais mais de 1.200 figuras públicas. Os autores reconhecem que os resultados de compromissos desse tipo foram mistos e que a falsa sinalização, em que a figura pública assina sem cumprir, pode impedir o compromisso de fornecer informação crível; levantam a hipótese de que quem assina sem outras formas de engajamento ativo terá impacto de longo prazo pequeno ou desprezível. O estudo empírico avaliou comportamento passado, e não o corrente, com 21 participantes recrutados por postagens no Facebook e e-mails a quem tinha assinado e pedido atualizações. Dois codificadores deram nota de 1 (a postagem é desinformação) a 5 (voltada a combater a desinformação e promover a verdade) às dez primeiras postagens com conteúdo noticioso feitas quatro semanas depois do compromisso, comparadas às dez primeiras do mesmo período do ano anterior. A média passou de 2,49 (DP = 0,60) para 3,65 (DP = 0,41), $d = -1,93$, com alfa de Krippendorff de 0,85, índice de concordância entre os dois codificadores.

Os autores escrevem que o estudo fornece evidência de que as pessoas se comportaram de modo mais honesto no Facebook porque assinaram o compromisso, e o descrevem como preliminar e voluntário, a replicar com amostra maior e grupos de controle. Acrescenta-se o que o desenho impõe: não há grupo de comparação; os 21 se autosselecionaram; a nota mistura a qualidade da postagem com o engajamento nos comentários; e a linguagem causal supera o que um pré-pós sem controle sustenta, porque quem decide assinar e pedir atualizações é quem está disposto a mudar. O que o estudo tem de aproveitável é a descrição da contingência arranjada, e não o efeito.

Os dois estudos agem sobre quem fala, depois ou antes do enunciado, e nenhum trata do que o Capítulo 3 identificou como o caminho da fala falsa em cadeia: o intraverbal com forma de tato. A seção 3.3 descreveu que a comunidade reforça a resposta com a forma de um relato direto do evento, ainda que o falante a emita a partir do que ouviu; a seção 3.6, que o autoclítico de fonte é o que diria que a resposta não é tato. Skinner (1957) nota que perguntar “Você viu ou alguém lhe contou?” é perguntar se a resposta foi tato, ecoico ou intraverbal ao comportamento verbal de outra pessoa; que em muitas comunidades é boa educação começar a falar com um autoclítico que indique de onde vem o que se diz; e que as contingências do comportamento autodescritivo são arranjadas pela comunidade quando ela tem razão para perguntar “O que você disse?”, “Você disse isso?” e “Por que disse isso?” (p. 314).

A hipótese aqui proposta leva a uma intervenção que não está em nenhum estudo lido: ensinar e reforçar, na comunidade, a discriminação entre tato e intraverbal com forma de tato, e a emissão do autoclítico de fonte. Tem dois braços. O do ouvinte consiste em perguntar pela fonte e não reforçar, com curta, repasse ou concordância, o enunciado que chega sem ela; a pergunta torna o intraverbal com forma de tato menos reforçado do que o enunciado com fonte marcada. O do falante consiste em marcar a fonte (“li em...”, “me disseram que...”) e em retratar-se, o que se aproxima do que o compromisso de Tsi-pursky et al. pede ao mandar verificar antes de compartilhar e retratar-se. O ponto de intervenção é o mesmo da seção 12.3, o momento da leitura, com uma diferença: o estímulo é uma consequência arranjada por outra pessoa, e não uma regra impressa. Um desenho no formato de Hangartner et al. serviria para testá-la: um perfil neutro que, diante de uma afirmação checada como falsa, pergunta de onde ela vem, contra um controle sem resposta, medindo a presença de marcadores de fonte nas postagens seguintes, a exclusão do enunciado e o repasse, em ambos os campos políticos.

<p>■ DADO</p>	<p>Entre 1.350 usuários do Twitter, a contra-fala por empatia aumentou a exclusão do tuíte em cerca de 8,4 pontos percentuais e reduziu, em 0,10 e 0,13 desvio-padrão, o número de tuítes xenófobos e o total, em quatro semanas, com valores de P ajustados de 0,053 a 0,112; a parcela de tuítes xenófobos, desfecho pré-registrado, não mudou, e humor e aviso não tiveram efeito consistente (Hangartner et al., 2021). Entre 21 signatários autosselecionados, a nota de postagens no Facebook passou de 2,49 para 3,65, sem grupo de controle (Tsipursky et al., 2018).</p>
<p>■ INTERPRETAÇÃO</p>	<p>Os dois estudos medem comportamento observado de quem foi alvo ou de quem assinou, com seguimento de quatro semanas, e nenhum mede o efeito sobre quem apenas lê. No estudo de campo, os dois componentes da parcela de tuítes xenófobos caíram em tamanho semelhante, em desvios-padrão, e o resultado é compatível com o usuário tuitar menos, e não com tuitar menos ódio. O estudo do compromisso é descrição de uma contingência com um pré-pós sem controle.</p>
<p>■ HIPÓTESE</p>	<p>Uma mensagem única de uma conta desconhecida pode funcionar como punição social leve, e então excluir seria escape, e tuitar menos, esquiva (leitura aqui proposta, ao lado da persuasão que os autores propõem). Nada nos dados separa as duas. Quanto ao autoclítico de fonte, a hipótese é a da seção: o ouvinte que pergunta pela fonte altera a consequência da forma de tato. Não há dado que a sustente nem que a refute.</p>

12.5 Reduzir preconceito por equivalência e por histórias em quadrinhos

O Capítulo 4 mostrou que as relações avaliativas podem ser formadas e reorganizadas em laboratório. A pergunta desta seção é outra: o que as fontes lidas dizem do uso desses procedimentos para reduzir preconceito.

Dixon e Lemke (2007) são o estudo que a literatura lida cita como modelo, e vale lê-lo pelo que fez. Participaram 12 estudantes de graduação (brancos ou afro-americanos) da Southern Illinois University, por crédito extra; os autores definem operacionalmente o preconceito como responder de modo diferencial a membros de uma raça específica. No pré-teste, cada participante avaliou 30 imagens (dez de homens brancos americanos, dez de homens do Oriente Médio e dez de objetos comuns), de 1 (mau) a 10 (bom). O treino foi de pareamento com o modelo, com consequência diferencial (som agradável para acerto, som aversivo para erro) e testes de simetria e equivalência em extinção; a relação treinada ligava a imagem de homem americano branco à pala-

vra “mau”, a de homem do Oriente Médio à palavra “bom” e a de objeto comum à palavra “neutro” (p. 10). Quatro participantes avaliaram todos os estímulos masculinos com “5” no pré-teste e tiveram os dados excluídos; os oito restantes avaliaram os homens do Oriente Médio mais perto de “mau” que os americanos.

Todos completaram o treino, e sete de oito atingiram 90% ou mais em cada teste de simetria e equivalência. Os autores escrevem que o homem do Oriente Médio passou a ser avaliado mais perto de “bom” e o americano mais perto de “mau”, e que o padrão se generalizou para nove estímulos novos. A Tabela 1 do artigo permite uma contagem que os autores não fazem: em cinco dos oito participantes (1, 3, 5, 8 e 12), a média para o homem do Oriente Médio ficou acima da do homem americano no pós-teste, e a média para o americano caiu para 1,00 a 1,50. Na discussão, dizem que nem todos os que atingiram o critério de equivalência mostraram diferenças significativas e atribuem parte da ausência de mudança a contingências sociais culturais implícitas, fonte de controle maior que a diferenciação por pontos usada no estudo; os objetos comuns, previstos como neutros, receberam as notas mais altas no pré e no pós-teste. Ressalvam que os participantes podem não responder conforme sua rede relacional verdadeira, por contingências sociais que reduzem a exibição de comportamento preconceituoso, e registram que, quando se demonstra equivalência entre estímulos com relações pré-existentes fortes, pode ser difícil alterar essas relações.

Dois pontos de leitura. O primeiro é o que o treino ensinou. Pela definição do Capítulo 4, uma atitude em que os atributos avaliativos são negativos é um preconceito; o procedimento de Dixon e Lemke relacionou “mau” ao homem americano branco e “bom” ao do Oriente Médio, de modo que o resultado esperado não é a neutralização, e sim a inversão do sinal, e a inversão instala em laboratório uma avaliação negativa de outro grupo. É leitura aqui proposta, e vale como cuidado de delineamento, e não como crítica aos autores, que o descrevem como demonstração. O segundo é a contagem: cinco de oito é uma maioria sem unanimidade, numa amostra pequena, sem grupo de controle e sem seguimento. Em crianças, com faces de pessoas negras e um símbolo positivo, e sem classe negativa para o outro grupo, o Capítulo 4 relata que 13 de 13 formaram as classes e nove mantiveram as relações (Mizael et al., 2016a), e que, em 46 crianças, 33 formaram classes e 27 mantiveram ao menos uma relação depois de seis semanas, com redução também no grupo controle, em menor grau (Mizael et al., 2021b). Mizael et al. (2016b) sustentam que um treino de uma ou duas horas é provavelmente insuficiente para modificar relações pré-experimentais amplamente reforçadas pela comunidade verbal; é hipótese dos revisores.

As histórias em quadrinhos entram aqui por serem o procedimento que os estudos do Instituto Walden⁴ usam, e por uma razão que interessa à intervenção: os autores de um dos volumes argumentam que as narrativas, por tratar o preconceito de forma sutil, diluída no enredo, minimizam a percepção de intenção persuasiva e permitem que o leitor se identifique com os personagens, o que os métodos tradicionais de mudança de atitude não fazem diante da resistência atitudinal (Moreira et al., 2022a). O procedimento, o elenco e as cenas estão nas Figuras 4.15 a 4.18 do Capítulo 4. O volume organizado por Moreira (2025a) reúne estudos de sujeito único, majoritariamente com universitários, com pareamento ao modelo e com quadrinhos, sobre sexismo, homofobia, etarismo, transfobia e gordofobia, entre outros, em que a atitude é medida por escala de diferencial semântico; o resumo relata que o pareamento formou classes e transferiu funções avaliativas, em especial para estímulos inicialmente neutros ou menos carregados emocionalmente, e que houve resistência à mudança de preconceitos profundamente enraizados ou associados a forte carga emocional. O Experimento 8 de Moreira et al. (2022c) fez cinco adultos lerem uma história com três personagens, de ideais políticos de esquerda, de direita e neutro, e avaliarem os personagens em três momentos; houve alteração nas escalas dos cinco, nos três personagens, num delineamento em que o sujeito é o próprio controle, em pré-teste e pós-teste. Aplica-se a esse estudo o critério aplicado aos outros: não há grupo de controle, não há seguimento, a medida é uma escala respondida logo depois e o objeto da avaliação é um personagem fictício, de modo que a alteração das marcações não se distingue, sozinha, de efeito de repetição da medida (Capítulo 4, seção 4.7).

Três revisões situam o conjunto. Matsuda et al. (2020) procuraram, em 14 periódicos analítico-comportamentais, por palavras-chave de racismo e preconceito no título (PsycInfo, desde 1990), 30 anos de literatura, e relatam que o resultado foi zero; ampliando a busca a artigos conceituais e a estudos funcional-contextuais, escrevem que há déficit notável da análise do comportamento sobre racismo e pouca pesquisa para orientar o trabalho. Resumem que os estudos de equivalência mostram que o viés existe e pode ser medido, e que vieses estabelecidos são difíceis de superar, embora haja procedimentos promissores; ressaltam que não afirmam que tais procedimentos mudem preconceitos a longo prazo nem que previnam comportamento racista explícito fora do laboratório. Citam, como relatos que não foram lidos no original, estudos de atenção plena e de tomada de perspectiva, um piloto de aula baseada em ACT e um estudo com professores, em que, depois de automonitoramento, feedback de desempenho e treino, crianças de todas as categorias receberam taxas equitativas de elogio e repreensão. Concluem que não apresentam solu-

ção definitiva, porque a pesquisa comportamental atual não é suficiente nem robusta, e recomendam estender e replicar estudos, estudar resultados longitudinais e desenvolver intervenções comunitárias.

Shea et al. (2023) revisam, em artigo narrativo sem método de busca descrito, as estratégias comuns contra o preconceito (educação, treino de diversidade e vergonha) pela Teoria das Molduras Relacionais, porque os resultados são limitados. Relatam, de trabalhos que não foram lidos no original, mudanças pequenas e autorrelatadas nas atitudes, que decaíram com o tempo, numa metanálise de 230 amostras em 40 anos; um treino de prevenção de assédio com pequeno efeito negativo nas atitudes de 90 universitários homens; uma literatura de treino de diversidade fragmentada e limitada por medidas pobres e falta de dados de manutenção e de comportamento; e um procedimento de vergonha sem efeito sobre atribuições relativas a deficiências. A hipótese dos autores é que o modo de entrega da educação pode fortalecer o preconceito ao reapresentar as visões preconceituosas, e que a vergonha pode funcionar como operação motivadora condicionada reflexiva que leva a esconder o preconceito. Concluem que as pesquisas em Teoria das Molduras Relacionais, em ACT e em IRAP mostraram resultados promissores, mas que é imprescindível pesquisa longitudinal, e que os tratamentos não devem perpetuar as redes relacionais preconceituosas; recomendam, como hipótese, treino de múltiplos exemplares com grupos marginalizados antes de as crianças serem expostas a redes que promovem preconceito.

Levy et al. (2022) acrescentam uma restrição de método. Dizem que a literatura analítico-comportamental muitas vezes deixa de relatar variáveis demográficas suficientes, inclusive de raça e de populações e contextos culturalmente diversos, o que pode levar a pouca compreensão da eficácia de medidas e avaliações. Os estudos desta seção, em geral, relatam pouco sobre quem são os participantes e para quem o resultado vale.

■ **DADO**

Em laboratório, o treino de equivalência alterou avaliações no sentido treinado em cinco de oito adultos (contagem própria sobre Dixon e Lemke, 2007), em 13 de 13 crianças com manutenção em nove (Mizael et al., 2016a) e em 33 de 46 com manutenção parcial em seis semanas (Mizael et al., 2021b). Uma busca em 14 periódicos não encontrou estudos com os termos de racismo ou preconceito no título (Matsuda et al., 2020), e as revisões dizem que não há dado de efeito de longo prazo nem de comportamento explícito fora do laboratório (Matsuda et al., 2020; Shea et al., 2023).

<p>■ INTERPRETAÇÃO</p>	<p>As relações avaliativas são treináveis e, em parte, mantêm-se por semanas, em amostras pequenas e com medida de escala. Nada do que foi lido mostra mudança no que as pessoas dizem ou fazem em outros contextos, nem no que a comunidade reforça.</p>
<p>■ HIPÓTESE</p>	<p>As relações antigas continuam reforçadas pela comunidade verbal fora do procedimento, e por isso o treino pontual tende a ser insuficiente, como sustentam Mizael et al. (2016b) e Shea et al. (2023) (hipóteses dos autores). Disso se tiram duas leituras: a de que a intervenção que muda só a relação no indivíduo, sem mexer no que a comunidade reforça, é a que as fontes menos sustentam; e a de que a forma da entrega (corrigir, instruir, envergonhar) pode reapresentar a relação que se quer desfazer, como a seção 12.1 levantou para a correção.</p>

12.6 Mudar as contingências da plataforma

Esta seção trata do nível da plataforma. Lazer et al. (2018) discutem duas estratégias gerais de mitigação: capacitar indivíduos a identificar notícias falsas e intervir nas plataformas para reduzir a atenção que recebem; e apontam que a cooperação das plataformas com pesquisadores é crítica para dimensionar o problema e a eficácia das intervenções. As seções anteriores trataram da primeira. Esta trata da segunda, que age sobre o que o Capítulo 8 descreveu como regra de entrega e o Capítulo 11 chamou de alcance. É também a que mais depende de acesso: os estudos lidos são poucos, e a maior parte foi feita com a cooperação da própria plataforma.

O que a regra de entrega faz foi medido por Huszár et al. (2022), no Twitter. Com a personalização da linha do tempo, introduzida em 2016, todos os partidos tiveram amplificação positiva (a razão entre o alcance no grupo com a linha do tempo personalizada e no grupo de controle, que seguiu a cronológica; 0% é alcance igual), em alguns casos acima de 200%, isto é, audiência três vezes a da linha do tempo cronológica; com exceção da Alemanha, o maior partido de direita ou centro-direita teve amplificação maior que o de esquerda ou centro-esquerda, e, no nível individual, um teste de permutação não detectou associação estatisticamente significativa entre o partido do político e sua amplificação. Não é uma intervenção avaliada, e sim a medida de uma regra que já estava em operação. O grupo de controle, de 1% dos usuários, não foi criado para pesquisa, e os próprios autores dizem que ele não é isolado de efeitos indiretos, de modo que as estatísticas provavelmente subestimam os efeitos; o trabalho foi revisado pelas equipes jurídica e de privacidade do Twitter, e não por comitê de ética.

Nyhan et al. (2023) fizeram o que Huszár et al. não puderam: uma intervenção randomizada no feed. Os 23.377 participantes, adultos dos Estados Unidos, foram recrutados por convites no alto do feed do Facebook, deram consentimento e responderam a pesquisas em ondas antes, durante e depois da eleição. Para os designados ao tratamento, todo o conteúdo de amigos, grupos e páginas que se previa compartilharem a inclinação política do participante foi rebaixado no feed, de 24 de setembro a 23 de dezembro de 2020, com a maior força de rebaixamento que um pré-teste indicou que não deixaria o feed quase vazio; o feed do grupo de controle não foi alterado sistematicamente. A exposição a fontes de mesma inclinação caiu de 53,7% para 36,2%; a de fontes cruzadas subiu de 20,7% para 27,9%, e a de fontes nem de mesma inclinação nem cruzadas, de 25,6% para 35,9%. Não houve efeito mensurável em oito medidas de atitude pré-registradas, como polarização afetiva, extremismo ideológico, avaliação de candidatos e crença em alegações falsas, e os autores descartam efeitos de $\pm 0,12$ desvio-padrão ou mais, isto é, de pouco mais de um décimo da variação típica de cada medida. Quem teve a exposição reduzida engajou-se mais, por exposição, com o conteúdo de mesma inclinação que ainda via. Entre as explicações possíveis do resultado nulo, os autores citam que o conteúdo político congenial é fração do que se vê, que o Facebook é fração da dieta de informação, que persuadir é difícil e que o desenho não capta efeitos de uso prévio nem efeitos acumulados ao longo de anos; observam ainda que reduzir a exposição poderia induzir as pessoas a buscar essa informação em outro lugar da internet.

O terceiro estudo trata da moderação. Álvarez-Benjumea e Winter (2018) pediram a 180 participantes, recrutados numa plataforma de crowdsourcing, que comentassem, num fórum simulado, nove fotos de temas controversos (como feminismo, direitos LGBT e refugiados), em quatro condições: base (dois comentários amistosos, dois neutros e dois hostis), censurada (conteúdo de ódio apagado), extremamente censurada (só comentários amistosos) e contra-fala (comentários hostis com réplicas que destacam a inaceitabilidade). Comentários hostis (nota 7 ou mais) foram 24 (5,88%) na base, 4 (0,99%) na condição censurada, 18 (4,44%) na contra-fala e 15 (4,10%) na extremamente censurada; só a redução da condição censurada foi significativa. Os autores sugerem que o comportamento normativo online é motivado por normas descritivas, o que os outros fazem, mais do que por normas injuntivas, e que as réplicas, sendo elas próprias hostis, podem acrescentar ambiguidade sobre a norma. Ressalvam que não têm dados sobre a troca entre censura e liberdade de expressão e que o anonimato do fórum impede processos de identificação de grupo.

Ceylan et al. (2023), que o Capítulo 9 leu, entram aqui por dois motivos. O primeiro é o estudo 4, que funciona como modelo experimental de uma regra de entrega: em 80 tentativas, os participantes ganhavam “+5 pontos” quando a escolha seguia o critério do treino (no de acurácia, compartilhar uma manchete verdadeira ou não compartilhar uma falsa; no de desinformação, o inverso) ou não tinham nenhum retorno (controle); no teste, sem retorno, o padrão persistiu. Depois do treino de acurácia, 66% das verdadeiras e 24% das falsas foram compartilhadas; no controle, 40% e 24%; depois do treino de desinformação, 54% e 33%. O que o retorno fez foi, no primeiro caso, aumentar o compartilhamento do verdadeiro sem reduzir o do falso, e, no segundo, aumentar o compartilhar em geral. O segundo motivo é a conclusão dos autores: os usuários habituais são parte integral dos modelos de lucro das redes baseados em anúncios e, por isso, as plataformas dificilmente criarão estruturas de recompensa que impeçam hábitos.

<p>■ DADO</p>	<p>A personalização da linha do tempo amplificou todos os partidos, com mais amplificação à direita que à esquerda nos maiores partidos de seis dos sete países e sem associação detectável no nível do político (Huszár et al., 2022). Rebaixar fontes de mesma inclinação no Facebook por três meses reduziu essa exposição de 53,7% para 36,2% e não teve efeito mensurável em oito medidas de atitude (Nyhan et al., 2023). Apagar comentários hostis num fórum simulado reduziu os comentários hostis de 5,88% para 0,99% (Álvarez-Benjumea & Winter, 2018).</p>
<p>■ INTERPRETAÇÃO</p>	<p>A regra de entrega muda o que as pessoas veem, e a moderação muda o que se publica num fórum simulado; mas, no único experimento randomizado do feed lido, uma redução de 17,5 pontos percentuais na exposição a fontes de mesma inclinação não mudou as atitudes medidas. Exposição e atitude são variáveis diferentes (Capítulo 11).</p>
<p>■ HIPÓTESE</p>	<p>O que a regra de entrega altera é a ocasião (o alcance) e, por meio da reação da audiência, a consequência de quem produz (Capítulo 9). Se for assim, a intervenção que alcança o que mantém a produção é a que muda o retorno que o produtor recebe, e nenhum estudo lido a testou: todos mediram o que a audiência vê ou faz, e nenhum, o que quem publica passou a publicar. Testá-la pediria variar o retorno (alcance ou reação visível) e medir o conteúdo publicado depois, com o mesmo critério para produtores de qualquer campo político.</p>

A Tabela 12.1 reúne as onze intervenções das seções 12.1 a 12.6, com o desenho, a medida, o resultado e o limite que o próprio estudo declara.

TABELA 12.1*As intervenções lidas: desenho, medida, resultado e limite*

Estudo	Intervenção e desenho	Medida	Resultado	Limite
Nyhan e Reifler (2010)	Correção embutida em notícia simulada; quatro experimentos entre sujeitos, estudantes	Concordância com a afirmação, logo depois da leitura	Backfire em dois de quatro, entre conservadores; a correção funcionou nos outros dois	Backfire só em conservadores; sem seguimento; amostra de estudantes
Freeze et al. (2021)	Aviso de desinformação, inválido ou válido; experimento online, 434 participantes	Credibilidade e memória	Aviso reduziu a credibilidade em 0,59 ponto (de 5), mesmo de artigo exato; aviso válido, 46% para 52% de memória ($p = 0,06$)	Aviso geral, do pesquisador; só a sessão
Roozenbeek e van der Linden (2019)	Jogo Bad News; pré-pós sem controle, 14.266 pares	Confiabilidade de 1 a 7	d de 0,30 a 0,36 em três técnicas; itens reais inalterados	Sem controle; autosselecionada, 75% homens; sem seguimento
Roozenbeek et al. (2022)	Vídeos de inoculação; seis experimentos randomizados e um de campo (22.632)	Reconhecimento da técnica, confiabilidade, disposição a compartilhar	Hipóteses sustentadas em 16 de 20 medidas; campo, $h = 0,09$	Até 24 horas; Estados Unidos; suscetibilidade, não comportamento
Guess et al. (2020)	Lista de dicas do Facebook; painel de duas ondas (EUA e Índia) e amostra presencial	Acurácia percebida, intenção de compartilhar, consumo observado	EUA: discriminação +26,5%; efeito cai em mais da metade em 20 dias; Índia presencial, sem efeito	Sem evidência em intenção e em consumo
Dixon e Lemke (2007)	Treino de equivalência com valência invertida; 12 estudantes, 8 analisados	Avaliação de 1 a 10 de imagens	Cinco de oito inverteram a ordem (contagem própria)	Sem controle nem seguimento; objetos com as notas mais altas
Mizael et al. (2021b)	Treino com parâmetros diferentes; 46 crianças; seguimento de seis semanas	Formação e manutenção de classes; viés	33 de 46 formaram classes; 27 mantiveram ao menos uma relação	Associação, não causa; controle também reduziu viés
Hangartner et al. (2021)	Contra-fala de aplicação única por perfil neutro; 1.350 usuários do Twitter; quatro semanas	Tuítes xenófobos criados e excluídos	Empatia: exclusão +0,21 DP (ajustado, 0,053); tuítes xenófobos -0,10 DP e total -0,13 DP (ajustado, 0,112); desfecho pré-registrado sem efeito	Não mede quem apenas lê; humor e aviso sem efeito

Estudo	Intervenção e desenho	Medida	Resultado	Limite
Tsipursky et al. (2018)	Compromisso público (Pro-Truth Pledge); pré-pós, 215, de dez postagens signatários	Nota de dois codificadores, de 1 a -1,93	2,49 para 3,65 (d = 0,99)	Sem controle; autosseleção; comportamento passado
Álvarez-Benjumea e Winter (2018)	Remoção de comentários hostis; (nota 7 ou mais) fórum simulado; 180 participantes	Comentários hostis	5,88% na base e 0,99% na censurada; contra-fala sem efeito	Sem dados sobre a troca com liberdade de expressão
Nyhan et al. (2023)	Rebaixar fontes de mesma inclinação no feed do Facebook; três meses; n = 23.377	Exposição e oito medidas de atitude	Exposição de 53,7% para 36,2%; sem efeito nas atitudes ($\pm 0,12$ DP excluído)	Efeitos acumulados não captados; Estados Unidos, 2020

Nota. Elaboração própria, a partir dos artigos; os números e as páginas estão nas seções 12.1 a 12.6. A coluna “Limite” reúne ressalvas dos autores e, em alguns casos, o que o desenho impõe e o texto discute; as contagens próprias vêm assinaladas.

12.7 Planejamento cultural e os limites éticos de quem planeja

Dos quatro níveis de intervenção do capítulo, o da cultura é o único em que nenhum estudo lido avaliou uma intervenção, a linha vazia da Figura 12.1. Não é por falta de conceito. Glenn (2004) trata, nas páginas finais de seu artigo, do que se pode planejar quando o objeto é uma prática e não uma pessoa, e é dela o conceito de que a análise precisa.

A leitura permite reordenar o que o capítulo viu. As intervenções das seções 12.1 a 12.3 e 12.5 agem sobre a ocasião de cada indivíduo, mostrando, ensinando ou avisando, e supõem, no vocabulário de Glenn, que mudar o comportamento de muitos pessoalmente mudaria o efeito cumulativo; as da seção 12.4 agem sobre a consequência que uma pessoa arranja para outra; as da 12.6 agem sobre o ambiente que seleciona o que circula. A hipótese dos Capítulos 8 e 9 é que o que mantém a fala está na consequência, e não na ocasião. Se for assim, a evidência está onde é barato medir, e as hipóteses estão onde está o que mantém. É a observação central do capítulo, e é interpretação: nada nos dados mostra que a intervenção sobre a consequência funcione melhor, porque nenhum estudo lido a testou.

Levy et al. (2022) dão ao planejamento uma condição que as outras fontes não dão. Descrevem a humildade cultural como valor que guia a interação com pessoas de outros contextos culturais, um padrão contínuo de comportamento, e não um resultado terminal alcançável, e criticam a “competência cultural” por sugerir que a aprendizagem tem um ponto final. Escrevem que os comportamentos que compõem o racismo são ensinados, imitados e manti-

dos, de modo incidental ou por contingências deliberadamente arranjadas, e que o racismo existe nos níveis individual, sistêmico e cultural. Mencionam as metacontingências selecionadas para manter sistemas racistas e dizem que examiná-las fica fora do escopo do artigo. E dizem que a análise do comportamento, com raízes eurocêntricas e falta observada de diversidade, equidade e inclusão, não está isenta de avaliação, e que precisa ser adaptada à variabilidade cultural de seus membros e das pessoas atendidas. Leitura aqui proposta: quem planeja é parte da prática que planeja, e a ciência que oferece as ferramentas também.

CONCEITO EM USO

Planejamento cultural (Glenn, 2004)

Glenn escreve que a engenharia cultural ainda não encontrou base científica segura, que o primeiro passo é entender os fenômenos a serem planejados, e que ela pode incidir sobre dois tipos de fenômeno: o macrocomportamento e as metacontingências. A única maneira de agir sobre os efeitos cumulativos do macrocomportamento é alterar o comportamento do maior número possível de participantes individuais; quando o número é grande, a mudança numa pequena porcentagem pode fazer diferença importante. Para desenhar intervenções, é preciso identificar as contingências operantes que explicam o comportamento dos indivíduos que participam da prática, o que implica modificar as contingências operantes que mantêm o comportamento de muitas pessoas. As contingências comportamentais entrelaçadas recorrentes podem ser mudadas de duas maneiras: alterando o ambiente selecionador externo e aguardando variações, ou alterando os componentes delas para que se adaptem melhor ao ambiente selecionador atual. Para Glenn, sem entender como as culturas surgem e evoluem, será difícil fazer escolhas sábias sobre o que pode ser mudado ou deve ser mudado. Leitura aqui proposta: planejamento cultural é o conjunto das intervenções que mudam contingências operantes de muitos participantes ou o ambiente que seleciona as práticas, e não a informação que cada um recebe.

Daí os limites éticos, que o capítulo reúne em cinco, todos extraídos das fontes e do critério aqui adotado. O primeiro é quem decide o que é falso ou o que se classifica como discurso de ódio. O Capítulo 2 mostrou que o rótulo de falso depende de quem verifica, com que evidência e em que momento, e o Capítulo 5 que “discurso de ódio” entra neste livro só como rótulo externo; planejar a redução de algo que se define por um terceiro faz do terceiro parte do planejamento. O segundo é a liberdade de expressão. Zielinski (2021) adverte que as ações corretivas precisam de salvaguardas de direitos humanos para não levarem à censura e à restrição da liberdade de expressão, e que a tecnologia sozinha não resolve; Hangartner et al. (2021) partem da premissa de que a mode-

ração pode suprimir fala valiosa e dispersar o ódio, e Álvarez-Benjumea e Winter (2018) dizem não ter dados sobre a troca entre censura e liberdade de expressão.

O terceiro limite é o do efeito que não é o buscado. A empatia de Hangartner et al. reduziu também o total de tuítes, e os autores pedem atenção ao arrefecimento para além do discurso de ódio; o aviso geral de Freeze et al. reduziu a credibilidade de notícias exatas; o treino de equivalência de Dixon e Lemke inverteu a valência; a correção de Nyhan e Reifler fez subir, entre conservadores, a concordância com a afirmação em dois dos quatro experimentos. Toda intervenção tem de ser lida também pelo que instala ao lado do que remove. O quarto é o do consentimento e do controle por terceiros. O trabalho de Huszár et al. foi revisado pelas equipas jurídica e de privacidade da empresa, no curso de suas operações ordinárias e não por um comitê de ética, e uma avaliação de impacto de proteção de dados concluiu que mecanismos adicionais de aviso e consentimento não eram necessários. O de Hangartner et al. foi aprovado por comitê de ética e pré-registrado, e a intervenção veio de perfis criados para o estudo, com histórico neutro e apolítico; o texto lido não traz menção a consentimento dos usuários. Não se decide aqui entre os dois modos; registra-se que, nos dois, a decisão sobre aviso e consentimento coube a quem conduziu ou revisou o estudo.

O quinto é o do critério simétrico. O que o Capítulo 10 estabeleceu para a janela de Overton vale aqui: o critério é o mesmo para o que alarga e para o que estreita, para o campo político de quem emite, para o idioma e para a plataforma, e a ressalva de que cada estudo mede o que mede é a mesma. Uma intervenção cujos efeitos só se medem para o conteúdo de um campo não passa no critério, qualquer que seja o campo.

Para quem quiser usar o capítulo na leitura de uma intervenção, de uma notícia sobre ela ou de um relatório de plataforma, as perguntas que as seções sugerem são as seguintes.

1. Em que nível age (indivíduo, interação, plataforma ou cultura) e sobre qual termo da contingência (ocasião, resposta ou consequência)?
2. O que se mediu: julgamento, intenção declarada ou comportamento observado? Com que seguimento, e em quem (quem lê, quem fala ou quem publica)?
3. Qual era a comparação, e a designação foi ao acaso? O que os autores dizem que o desenho não permite afirmar?
4. O resultado vale para conteúdo de outro campo político, de outro idioma e de outra plataforma? O critério foi o mesmo para todos?

5. O que a intervenção instala ou ensina além do alvo: um sinal sem alvo, uma valência invertida, uma regra que compete com a reação da audiência?
6. Quem decide o que conta como falso ou como discurso de ódio, e que salvaguarda há contra a restrição da expressão?

Síntese

O capítulo organizou as intervenções em quatro níveis e leu onze estudos pelo que cada um mediu. Oito mediram julgamento ou desempenho em tarefa, cinco mediram comportamento observado, quatro, comportamento com seguimento de uma semana ou mais, e nenhum avaliou uma intervenção em nível cultural (Figura 12.1). A correção depois da leitura tem efeito incerto e depende do tema; o backfire de Nyhan e Reifler veio de percepções errôneas de conservadores, nos temas escolhidos, e o aviso geral reduziu a credibilidade de notícias exatas (12.1).

A inoculação, lida como treino discriminativo com exemplares da técnica, melhora, logo depois, a avaliação de itens novos que a usam, num efeito pequeno no campo e sem medida de comportamento (12.2). As dicas de identificação de notícias falsas mudaram o julgamento por pouco tempo, onde foram entendidas, e não mudaram, nos dados, o consumo nem, com evidência suficiente, a intenção declarada de compartilhar (12.3). Diante de quem fala, a contra-fala por empatia aumentou a exclusão e reduziu a atividade, sem efeito no desfecho pré-registrado, e o compromisso público tem um pré-pós sem controle (12.4); a hipótese aqui proposta, a de ensinar e reforçar o autoclítico de fonte, não tem teste.

O treino de equivalência altera avaliações em laboratório, em parte e por semanas, e nenhuma fonte lida mostra mudança do que se diz ou faz fora dele; as revisões pedem seguimento e intervenção comunitária, e uma delas ressalva que reapresentar a relação pode fortalecê-la (12.5). A regra de entrega muda o que se vê, e rebaixar fontes de mesma inclinação não mudou as atitudes medidas; a moderação reduziu os comentários hostis num fórum simulado, e nenhum estudo mediu o que quem publica passou a publicar (12.6). Para Glenn, agir sobre efeitos cumulativos pede mudar as contingências operantes de muitos participantes ou o ambiente que seleciona as práticas; os limites éticos são o de quem define o que é falso, a liberdade de expressão, o efeito colateral, o consentimento e a simetria do critério (12.7).

O que o capítulo deixa é uma assimetria. A evidência está onde é barato medir, no julgamento de uma pessoa diante de uma tela, e as hipóteses estão onde, segundo a análise dos Capítulos 8 e 9, está o que mantém a fala: na consequên-

cia que a audiência e a plataforma arranjam para quem a produz. A pergunta que fica para quem estuda é a de um desenho que meça o comportamento de quem publica, diante de uma mudança de retorno, com seguimento e com o mesmo critério para todos os campos.

Referências

- Álvarez-Benjumea, A., & Winter, F. (2018). Normative change and culture of hate: An experiment in online environments. *European Sociological Review*, 34(3), 223–237. <https://doi.org/10.1093/esr/jcy005>
- Calleja, N., AbdAllah, A., Abad, N., Ahmed, N., Albarracin, D., Altieri, E., Anoko, J. N., Arcos, R., Azlan, A. A., Bayer, J., Bechmann, A., Bezbaruah, S., Briand, S. C., Brooks, I., Bucci, L. M., Burzo, S., Czerniak, C., De Domenico, M., Dunn, A. G., . . . Purnat, T. D. (2021). A public health research agenda for managing infodemics: Methods and results of the first WHO infodemiology conference. *JMIR Infodemiology*, 1(1), Article e30979. <https://doi.org/10.2196/30979>
- Ceylan, G., Anderson, I. A., & Wood, W. (2023). Sharing of misinformation is habitual, not just lazy or biased. *Proceedings of the National Academy of Sciences*, 120(4), Article e2216614120. <https://doi.org/10.1073/pnas.2216614120>
- Dixon, M. R., & Lemke, M. (2007). Reducing prejudice towards Middle Eastern persons as terrorists. *European Journal of Behavior Analysis*, 8(1), 5–12. <https://doi.org/10.1080/15021149.2007.11434269>
- Dixon, M. R., Dymond, S., Rehfeldt, R. A., Roche, B., & Zlomke, K. R. (2003). Terrorism and relational frame theory. *Behavior and Social Issues*, 12(2), 129–147. <https://doi.org/10.5210/bsi.v12i2.40>
- Freeze, M., Baumgartner, M., Bruno, P., Gunderson, J. R., Olin, J., Ross, M. Q., & Szafran, J. (2021). Fake claims of fake news: Political misinformation, warnings, and the tainted truth effect. *Political Behavior*, 43(4), 1433–1465. <https://doi.org/10.1007/s11109-020-09597-3>
- Glenn, S. S. (2004). Individual behavior, culture, and social change. *The Behavior Analyst*, 27(2), 133–151. <https://doi.org/10.1007/bf03393175>
- Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences*, 117(27), 15536–15545. <https://doi.org/10.1073/pnas.1920498117>
- Hangartner, D., Gennaro, G., Alasiri, S., Bahrich, N., Bornhoft, A., Boucher, J., Demirci, B., Derksen, L., Hall, A., Jochum, M., Munoz, M. M., Richter, M., Vogel, F., Wittwer, S., Wüthrich, F., Gilardi, F., & Donnay, K. (2021). Empathy-based counterspeech can reduce racist hate speech in a social media field experiment. *Proceedings of the National Academy of Sciences*, 118(50), Article e2116310118. <https://doi.org/10.1073/pnas.2116310118>
- Huszár, F., Ktena, S. I., O'Brien, C., Belli, L., Schlaikjer, A., & Hardt, M. (2022). Algorithmic amplification of politics on Twitter. *Proceedings of the National Academy of Sciences*, 119(1), Article e2025334119. <https://doi.org/10.1073/pnas.2025334119>

- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094–1096. <https://doi.org/10.1126/science.aao2998>
- Levy, S., Siebold, A., Vaidya, J., Truchon, M.-M., Dettmering, J., & Mittelman, C. (2022). A look in the mirror: How the field of behavior analysis can become anti-racist. *Behavior Analysis in Practice*, 15(4), 1112–1125. <https://doi.org/10.1007/s40617-021-00630-3>
- Matsuda, K., Garcia, Y., Catagnus, R., & Brandt, J. A. (2020). Can behavior analysis help us understand and reduce racism? A review of the current literature. *Behavior Analysis in Practice*, 13(2), 336–347. <https://doi.org/10.1007/s40617-020-00411-4>
- Mizael, T. M., de Almeida, J. H., Roche, B., & de Rose, J. C. (2021b). Effectiveness of different training and testing parameters on the formation and maintenance of equivalence classes: Investigating prejudiced racial attitudes. *The Psychological Record*, 71(2), 265–277. <https://doi.org/10.1007/s40732-020-00435-w>
- Mizael, T. M., de Almeida, J. H., Silveira, C. C., & de Rose, J. C. (2016a). Changing racial bias by transfer of functions in equivalence classes. *The Psychological Record*, 66(3), 451–462. <https://doi.org/10.1007/s40732-016-0185-0>
- Mizael, T. M., Dos Santos, S. L., & De Rose, J. C. C. (2016b). Contribuições do paradigma de equivalência de estímulos para o estudo das atitudes. *Interação em Psicologia*, 20(2). <https://doi.org/10.5380/psi.v20i2.46278>
- Moreira, M. B. (Org.). (2025a). *Formação e mudança de atitudes: Pareamento ao modelo e história em quadrinhos* (1ª ed.). Instituto Walden4.
- Moreira, M. B., Premuli, I. T., Rabelo, E. de P., Pessoa, M. B. F., & Medeiros, M. R. G. C. (2022a). *Atitudes: Histórias em quadrinhos, análises semânticas em mídias sociais e o paradigma de relações de equivalência de estímulos* (1ª ed.). Instituto Walden4. https://www.walden4.com.br/livros/pdfs/moreira_etal_2022_q0.pdf
- Moreira, M. B., Sousa, C. P., Tomasini, C., Craveiro, H. do A., Rodrigues, I. L., Dutra, I., Barros, J. F., Freire, L. R. V., D'arcanchy e Vilela, M. E., Cautela, M. E. T. C., Oliveira, P. M. D. N., Soares, S. K. S., & Andrade, V. M. (2022c). *Estudos sobre histórias em quadrinhos e atitudes - Parte I* (1ª ed.). Instituto Walden4. https://www.walden4.com.br/livros/pdfs/moreira_etal_2022_hq1.pdf
- Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303–330. <https://doi.org/10.1007/s11109-010-9112-2>
- Nyhan, B., Settle, J., Thorson, E., Wojcieszak, M., Barberá, P., Chen, A. Y., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Guess, A. M., Kennedy, E., Kim, Y. M., Lazer, D., Malhotra, N., Moehler, D., . . . Tucker, J. A. (2023). Like-minded sources on Facebook are prevalent but not polarizing. *Nature*, 620(7972), 137–144. <https://doi.org/10.1038/s41586-023-06297-w>
- Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, 5(1), Article 65. <https://doi.org/10.1057/s41599-019-0279-9>
- Roozenbeek, J., van der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances*, 8(34), Article eabo6254. <https://doi.org/10.1126/sciadv.abo6254>

- Shea, P., Johnson, P., & Togade, D. (2023). Using relational frame theory to examine racial prejudice: A tool for educators and an appeal for future research. *Behavior Analysis in Practice*, 16(1), 102–116. <https://doi.org/10.1007/s40617-022-00767-9>
- Skinner, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts.
- Tsipursky, G., Votta, F., & Roose, K. M. (2018). Fighting fake news and post-truth politics with behavioral science: The Pro-Truth Pledge. *Behavior and Social Issues*, 27(1), 47–70. <https://doi.org/10.5210/bsi.v27i0.9127>
- Zielinski, C. (2021). Infodemics and infodemiology: A short history, a long future. *Revista Panamericana de Salud Pública*, 45, 1. <https://doi.org/10.26633/rpsp.2021.40>

A

Ficha de análise de discurso

A ficha é a do Capítulo 3 (seção 3.2), pronta para uso. Ela existe para obrigar a separar o que se sabe do que se infere: os cinco primeiros campos são dado, os três seguintes são interpretação e os dois últimos são hipótese. Onde a fonte não documenta, escreve-se “não documentado”; o campo não se preenche por inferência. A mesma ficha, com o mesmo critério, vale para falas de qualquer campo político.

#	Campo	O que registrar	Natureza
1	Fala	O enunciado, transcrito como a fonte o traz, entre aspas, com a referência da fonte	Dado
2	Data	Data da fala	Dado
3	Contexto e veículo	Onde e em que veículo foi dita; a pergunta, o argumento ou o evento imediatamente anterior, se a fonte os traz	Dado
4	Audiência presente	Quem estava presente e como reagiu, quando a fonte documenta	Dado
5	Checagem	Agência, rótulo, data e justificativa; a falsidade é sempre da agência que a classificou (Capítulo 2)	Dado
6	Operante provável	Mando, tato, intraverbal, tato distorcido, Interpretação mando disfarçado ou tato impuro, função ritual ou de identidade; e por quê (Tabelas 3.1 e 3.4)	Interpretação
7	Variáveis de controle candidatas	Ocasão, audiência, consequência provável, história de reforço, o que o ouvinte faria depois	Interpretação
8	Marcas de forma	Autoclíticos de fonte e de certeza, quantificadores, repetição (pura ou com variação)	Interpretação
9	O que mudaria a fala	A observação que decidiria entre as leituras: a ocasião, a audiência, a comparação entre audiências ou antes e depois da correção	Hipótese

#	Campo	O que registrar	Natureza
10	Interpretações rivais	As seis da Tabela 3.4, com o que separaria cada uma: controle fraco pelo evento, intraverbal adquirido de outros, tato distorcido, mando disfarçado ou tato impuro, função ritual ou de identidade, efeito da seleção pela agência	Hipótese

Quatro tipos de observação a ficha pede e que registros feitos por terceiros costumam não trazer: a ocasião (o que a fala responde); a audiência (quem estava presente e como reagiu); a comparação do mesmo falante e do mesmo evento perante audiências diferentes, ou antes e depois da checagem; e o que acontece depois da correção (a forma permanece, varia ou desaparece). Sem elas, a análise permanece interpretação.

Ficha em branco

Campo	Registro
1. Fala	
2. Data	
3. Contexto e veículo	
4. Audiência presente	
5. Checagem (agência, rótulo, data)	
6. Operante provável	
7. Variáveis de controle candidatas	
8. Marcas de forma	
9. O que mudaria a fala e como se saberia	
10. Interpretações rivais e o que as separa	
Não documentado	

B

Conceitos em uso

Cada conceito é apresentado uma vez, num quadro “Conceito em uso” (ou “Definição proposta”, quando o termo é proposto aqui), no primeiro capítulo cuja análise depende dele; os capítulos seguintes o retomam e remetem ao quadro original. A tabela segue a ordem do livro.

Conceito	Capítulo (seção)	Definição curta	Fonte principal
Comportamento verbal, audiência e autorrelato	1 (1.5)	Comportamento reforçado pela mediação de outras pessoas; a audiência é o estímulo discriminativo em cuja presença ele é caracteristicamente reforçado. Autorrelato (escala, pergunta, entrevista) é comportamento verbal, sob controle de quem pergunta e de como pergunta	Skinner (1957); Guerin (1994)
Crença e atitude como comportamento verbal	2 (2.8)	Atitudes e crenças são comportamento moldado por comunidades verbais, e não relatos de estado interno; a crença pode ser contestada com evidência, a atitude não do mesmo modo	Guerin (1994)
Autoclítico	2 (2.8); 3 (3.6)	Qualificação acrescentada ao enunciado que modifica o efeito dele sobre o ouvinte	Guerin (1994); Skinner (1957)
Tríplice contingência, função × topografia e episódio verbal	3 (3.1)	O operante verbal se define pelo que controla a resposta, e não pela forma; a unidade é ocasião, resposta e consequência; o episódio inclui falante e ouvinte	Skinner (1957)

Conceito	Capítulo (seção)	Definição curta	Fonte principal
Reforço social generalizado	3 (3.3); retomado em 8 (8.1) e 9 (9.1)	Evento que costuma preceder muitos reforçadores diferentes e reforça o comportamento sob várias condições de privação; a aprovação é o exemplo. Para Guerin, atitudes são enunciados mantidos por ele	Skinner (1957); Guerin (1994)
Tato distorcido e subcomunidade verbal	3 (3.4); retomado em 6 (6.1)	Tato cujo controle pelo evento foi deslocado por consequências especiais do ouvinte; uma subcomunidade menor que a ampla pode controlar o que se relata	Skinner (1957); Guerin (1992)
Autoclíticos de fonte e de certeza	3 (3.6)	Autoclíticos que informam a fonte ou a força da resposta; quantificadores (“todos”, “nenhum”) modificam a reação do ouvinte à sentença inteira	Skinner (1957)
Regra e comportamento governado por regras	3 (3.7)	Estímulo discriminativo verbal que descreve uma contingência; o desempenho instruído é menos sensível a mudanças nas consequências	Baum (1995); Cerutti (1989); Hayes et al. (1986)
Atitude	4 (4.1)	Resposta verbal de atribuir a um objeto um valor de positivo a negativo, mantida pelas consequências que a comunidade verbal arranja; o objeto não precisa ter sido contatado	Guerin (1994); Cameoka e Moreira (2021)
Discriminação condicional e pareamento ao modelo	4 (4.2)	Dado um modelo e comparações, a escolha é conseqüenciada conforme o modelo: se A1, então B1	Sidman e Tailby (1982)
Classe de equivalência	4 (4.2)	Conjunto de estímulos cujas relações condicionais têm reflexividade, simetria e transitividade	Sidman e Tailby (1982)
Transferência (ou transformação) de função	4 (4.3)	Função adquirida por contato direto por um estímulo aparece nos demais membros da classe sem treino adicional; que tenha “transferido” é interpretação	Dougher et al. (1994); Dymond e Rehfeldt (2000)
Relações derivadas de oposição e de comparação	4 (4.6)	Padrões de responder relacional sob controle contextual: na oposição, A oposto de Y e de Z implica Y igual a Z; na comparação, a relação define como et a função se transforma	Dixon et al. (2003); Weinstein et al. (2008)
Contingências comportamentais entrelaçadas	5 (5.1); retomadas em 9 (9.4)	O comportamento de duas ou mais pessoas é evento ambiental para o das outras; o resultado é o do comportamento inter-relacionado	Glenn (2004)
Mando disfarçado	5 (5.2)	Enunciado com forma de tato ou de opinião cuja função, na evidência disponível, é a de mando; a aplicação à fala política é hipótese aqui proposta	Skinner (1957); Guerin (1994)

Conceito	Capítulo (seção)	Definição curta	Fonte principal
Controle aversivo	5 (5.3)	Controle do comportamento por estimulação aversiva (escape, esquiva, punição); pode ser verbal e branda	Skinner (1957); Fazzano e Gallo (2015)
Contracontrole	6 (6.4)	Episódios de condições controladoras aversivas socialmente mediadas e de respostas de fuga ou esquiva que não reforçam o comportamento controlador	Delprato (2002)
Esquiva de contracontrole	6 (6.4)	Termo proposto aqui, como hipótese: repertório do falante ou da comunidade mantido por reduzir a probabilidade ou a eficácia do contracontrole	Proposta aqui feita, a partir de Delprato (2002)
Correspondência dizer-fazer	7 (7.4)	Relação entre o que a pessoa diz que fez ou fará e o que faz; são classes separadas, e a correspondência é efeito de contingência	Risley e Hart (1968); Lloyd (1994)
Atenção como reforçador condicionado generalizado	8 (8.1)	A atenção do ouvinte é consequência comum de mandos generalizados e reforça o comportamento verbal sem respeito à forma	Skinner (1957)
Reforçamento intermitente	8 (8.4)	Só parte das respostas ou das ocasiões é seguida do reforçador; no comportamento verbal, a intermitência é a regra	Skinner (1957)
Esquemas de reforçamento	8 (8.4)	O reforçador depende do tempo desde o anterior (intervalo) ou do número de respostas (razão), fixos ou variáveis	Ferster e Skinner (1957)
Metacontingência e macrocontingência	9 (9.7); macrocontingência retomada em 11 (11.4)	Metacontingência: prática cultural e resultado agregado das variações correntes, que seleciona contingências entrelaçadas. Macrocontingência: relação entre a prática e a soma agregada das consequências, que não controla o indivíduo	Glenn (1988, 2004); Houmanfar e Rodrigues (2006)
Audiência negativa e audiência não punitiva	10 (10.1)	Audiência negativa: em sua presença o comportamento verbal é punido; não punitiva: apenas não reforça. Duas audiências juntas geram dilema	Skinner (1957)
Janela de uma audiência	10 (10.1)	Definição proposta: para um falante e um assunto, os enunciados que ele emite diante dessa audiência sem que ela os puna	Proposta aqui feita
Alcance, exposição e reação	11 (11.2)	Definição proposta: alcance é quem recebe ou pode ver a emissão; exposição, o que efetivamente viu; reação, o que faz depois	Proposta aqui feita

Conceito	Capítulo (seção)	Definição curta	Fonte principal
Treino discriminativo	12 (12.2)	O estímulo prévio passa a ser a ocasião da resposta por discriminação (1957); operante; com múltiplos exemplares, a resposta fica sob controle da propriedade comum e generaliza	Skinner et al. (2023)
Planejamento cultural	12 (12.7)	Intervenções que mudam as contingências operantes de muitos participantes ou o ambiente que seleciona as práticas (macrocomportamento e metacontingências)	Glenn (2004)



Referências

As obras citadas nos capítulos, em ordem alfabética (APA 7). Quando autores e ano coincidem, a letra que acompanha o ano é a das citações do texto. Todas foram lidas no texto integral, na versão indicada no Apêndice D (seção D.3).

- Agência Lupa. (2026, 27 de julho). *Anatomia da desinformação eleitoral* [Relatório]. https://uploads.agencialupa.org/2026/07/RELATORIO-ANATOMIA-DA-DESINFORMACAO-ELEITORAL_LUPA-FINAL.pdf
- Ajzen, I. (2001). Nature and operation of attitudes. *Annual Review of Psychology*, 52, 27–58. <https://doi.org/10.1146/annurev.psych.52.1.27>
- Ajzenman, N., Cavalcanti, T., & Da Mata, D. (2023). More than words: Leaders' speech and risky behavior during a pandemic. *American Economic Journal: Economic Policy*, 15(3), 351–371. <https://doi.org/10.1257/pol.20210284>
- Albarracin, D., & Shavitt, S. (2018). Attitudes and attitude change. *Annual Review of Psychology*, 69, 299–327. <https://doi.org/10.1146/annurev-psych-122216-011911>
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211–236. <https://doi.org/10.1257/jep.31.2.211>
- Álvarez-Benjumea, A., & Winter, F. (2018). Normative change and culture of hate: An experiment in online environments. *European Sociological Review*, 34(3), 223–237. <https://doi.org/10.1093/esr/jcy005>
- Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B. F., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37), 9216–9221. <https://doi.org/10.1073/pnas.1804840115>
- Barnes, D., Lawlor, H., Smeets, P. M., & Roche, B. (1996). Stimulus equivalence and academic self-concept among mildly mentally handicapped and nonhandicapped children. *The Psychological Record*, 46(1), 87–107. <https://doi.org/10.1007/bf03395165>
- Baum, W. M. (1995). Rules, culture, and fitness. *The Behavior Analyst*, 18(1), 1–21. <https://doi.org/10.1007/bf03392688>
- Bohner, G., & Dickel, N. (2011). Attitudes and attitude change. *Annual Review of Psychology*, 62, 391–417. <https://doi.org/10.1146/annurev.psych.121208.131609>
- Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., & Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415), 295–298. <https://doi.org/10.1038/nature11421>

- Bonow Soares, F., Recuero, R., Volcan, T., Fagundes, G., & Sodré, G. (2021). Research note: Bolsonaro's firehose: How Covid-19 disinformation on WhatsApp was used to fight a government political crisis in Brazil. *Harvard Kennedy School Misinformation Review*, 2(1). <https://doi.org/10.37016/mr-2020-54>
- Bortoloti, R., & de Rose, J. C. (2009). Assessment of the relatedness of equivalent stimuli through a semantic differential. *The Psychological Record*, 59(4), 563–590. <https://doi.org/10.1007/bf03395682>
- Brady, W. J., McLoughlin, K., Doan, T. N., & Crockett, M. J. (2021). How social learning amplifies moral outrage expression in online social networks. *Science Advances*, 7(33), Article eabe5641. <https://doi.org/10.1126/sciadv.abe5641>
- Bursztyn, L., Egorov, G., & Fiorin, S. (2020). From extreme to mainstream: The erosion of social norms. *American Economic Review*, 110(11), 3522–3548. <https://doi.org/10.1257/aer.20171175>
- Calleja, N., AbdAllah, A., Abad, N., Ahmed, N., Albarracin, D., Altieri, E., Anoko, J. N., Arcos, R., Azlan, A. A., Bayer, J., Bechmann, A., Bezbaruah, S., Briand, S. C., Brooks, I., Bucci, L. M., Burzo, S., Czerniak, C., De Domenico, M., Dunn, A. G., . . . Purnat, T. D. (2021). A public health research agenda for managing infodemics: Methods and results of the first WHO infodemiology conference. *JMIR Infodemiology*, 1(1), Article e30979. <https://doi.org/10.2196/30979>
- Cameoka, M. C., & Moreira, M. B. (2021). Preconceito racial: viés na mensuração de atitudes produzido por controle de estímulos. *Acta Comportamentalia*, 29(1), 93–112. <https://doi.org/10.32870/ac.v29i1.78782>
- Carvalho, P. R., Sousa, P. C. C., & Schneider, M. A. F. (2021). Desinformação na pandemia de Covid-19: similitudes informacionais entre Trump e Bolsonaro. *Em Questão*, 27(3), 15–41. <https://doi.org/10.19132/1808-5245273.15-41>
- Catania, A. C., Matthews, B. A., & Shimoff, E. (1982). Instructed versus shaped human verbal behavior: Interactions with nonverbal responding. *Journal of the Experimental Analysis of Behavior*, 38(3), 233–248. <https://doi.org/10.1901/jeab.1982.38-233>
- Cerutti, D. T. (1989). Discrimination theory of rule-governed behavior. *Journal of the Experimental Analysis of Behavior*, 51(2), 259–276. <https://doi.org/10.1901/jeab.1989.51-259>
- Ceylan, G., Anderson, I. A., & Wood, W. (2023). Sharing of misinformation is habitual, not just lazy or biased. *Proceedings of the National Academy of Sciences*, 120(4), Article e2216614120. <https://doi.org/10.1073/pnas.2216614120>
- Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), Article e2023301118. <https://doi.org/10.1073/pnas.2023301118>
- Cinelli, M., Quattrociocchi, W., Galeazzi, A., Valensise, C. M., Brugnoli, E., Schmidt, A. L., Zola, P., Zollo, F., & Scala, A. (2020). The COVID-19 social media infodemic. *Scientific Reports*, 10(1), Article 16598. <https://doi.org/10.1038/s41598-020-73510-5>
- de Carvalho, M. P., & de Rose, J. C. (2014). Understanding racial attitudes through the stimulus equivalence paradigm. *The Psychological Record*, 64(3), 527–536. <https://doi.org/10.1007/s40732-014-0049-4>
- de Rose, J. C., McIlvane, W. J., Dube, W. V., Galpin, V. C., & Stoddard, L. T. (1988). Emergent simple discrimination established by indirect relation to differential consequences. *Journal of the Experimental Analysis of Behavior*, 50(1), 1–20. <https://doi.org/10.1901/jeab.1988.50-1>

- DeFleur, M. L., & Westie, F. R. (1958). Verbal attitudes and overt acts: An experiment on the salience of attitudes. *American Sociological Review*, 23(6), 667–673. <https://doi.org/10.2307/2089055>
- Delprato, D. J. (2002). Countercontrol in behavior analysis. *The Behavior Analyst*, 25(2), 191–200. <https://doi.org/10.1007/bf03392057>
- Dixon, M. R., & Lemke, M. (2007). Reducing prejudice towards Middle Eastern persons as terrorists. *European Journal of Behavior Analysis*, 8(1), 5–12. <https://doi.org/10.1080/15021149.2007.11434269>
- Dixon, M. R., Dymond, S., Rehfeldt, R. A., Roche, B., & Zlomke, K. R. (2003). Terrorism and relational frame theory. *Behavior and Social Issues*, 12(2), 129–147. <https://doi.org/10.5210/bsi.v12i2.40>
- Dougher, M. J., Augustson, E., Markham, M. R., Greenway, D. E., & Wulfert, E. (1994). The transfer of respondent eliciting and extinction functions through stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior*, 62(3), 331–351. <https://doi.org/10.1901/jeab.1994.62-331>
- Dymond, S., & Rehfeldt, R. A. (2000). Understanding complex behavior: The transformation of stimulus functions. *The Behavior Analyst*, 23(2), 239–254. <https://doi.org/10.1007/bf03392013>
- Effron, D. A., & Raj, M. (2020). Misinformation and morality: Encountering fake-news headlines makes them seem less unethical to publish and share. *Psychological Science*, 31(1), 75–87. <https://doi.org/10.1177/0956797619887896>
- Evangelista, R., & Bruno, F. (2019). WhatsApp and political instability in Brazil: Targeted messages and political radicalisation. *Internet Policy Review*, 8(4). <https://doi.org/10.14763/2019.4.1434>
- Eysenbach, G. (2009). Infodemiology and infoveillance: Framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. *Journal of Medical Internet Research*, 11(1), Article e11. <https://doi.org/10.2196/jmir.1157>
- Eysenbach, G. (2020). How to fight an infodemic: The four pillars of infodemic management. *Journal of Medical Internet Research*, 22(6), Article e21820. <https://doi.org/10.2196/21820>
- Fazzano, L. H., & Gallo, A. E. (2015). Uma análise da homofobia sob a perspectiva da análise do comportamento. *Temas em Psicologia*, 23(3), 535–545. <https://doi.org/10.9788/tp2015.3-02>
- Ferster, C. B., & Skinner, B. F. (1957). *Schedules of reinforcement*. Appleton-Century-Crofts. <https://doi.org/10.1037/10627-000>
- Freeze, M., Baumgartner, M., Bruno, P., Gunderson, J. R., Olin, J., Ross, M. Q., & Szafran, J. (2021). Fake claims of fake news: Political misinformation, warnings, and the tainted truth effect. *Political Behavior*, 43(4), 1433–1465. <https://doi.org/10.1007/s11109-020-09597-3>
- Frimer, J. A., Aujla, H., Feinberg, M., Skitka, L. J., Aquino, K., Eichstaedt, J. C., & Willer, R. (2023). Incivility is rising among American politicians on Twitter. *Social Psychological and Personality Science*, 14(2), 259–269. <https://doi.org/10.1177/19485506221083811>
- Galizio, M. (1979). Contingency-shaped and rule-governed behavior: Instructional control of human loss avoidance. *Journal of the Experimental Analysis of Behavior*, 31(1), 53–70. <https://doi.org/10.1901/jeab.1979.31-53>

- Garrett, R. K., & Bond, R. M. (2021). Conservatives' susceptibility to political misperceptions. *Science Advances*, 7(23), Article eabf1234. <https://doi.org/10.1126/sciadv.abf1234>
- Glenn, S. S. (1988). Contingencies and metacontingencies: Toward a synthesis of behavior analysis and cultural materialism. *The Behavior Analyst*, 11(2), 161–179. <https://doi.org/10.1007/bf03392470>
- Glenn, S. S. (2004). Individual behavior, culture, and social change. *The Behavior Analyst*, 27(2), 133–151. <https://doi.org/10.1007/bf03393175>
- Goel, S., Anderson, A., Hofman, J., & Watts, D. J. (2016). The structural virality of online diffusion. *Management Science*, 62(1), 180–196. <https://doi.org/10.1287/mnsc.2015.2158>
- Guerin, B. (1992). Behavior analysis and the social construction of knowledge. *American Psychologist*, 47(11), 1423–1432. <https://doi.org/10.1037/0003-066x.47.11.1423>
- Guerin, B. (1994). Attitudes and beliefs as verbal behavior. *The Behavior Analyst*, 17(1), 155–163. <https://doi.org/10.1007/bf03392661>
- Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. *Proceedings of the National Academy of Sciences*, 117(27), 15536–15545. <https://doi.org/10.1073/pnas.1920498117>
- Guess, A., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science Advances*, 5(1), Article eaau4586. <https://doi.org/10.1126/sciadv.aau4586>
- Gunnarsson, K. F., Rzeszutek, M. J., Malkin, A., Tewogbola, P. O., Han, S., & Jacobs, E. A. (2026). A function-based examination of social media demand. *Behavior Analysis in Practice*. Advance online publication. <https://doi.org/10.1007/s40617-026-01197-7>
- Hangartner, D., Gennaro, G., Alasiri, S., Bahrich, N., Bornhoft, A., Boucher, J., Demirci, B. B., Derksen, L., Hall, A., Jochum, M., Munoz, M. M., Richter, M., Vogel, F., Wittwer, S., Wüthrich, F., Gilardi, F., & Donnay, K. (2021). Empathy-based counterspeech can reduce racist hate speech in a social media field experiment. *Proceedings of the National Academy of Sciences*, 118(50), Article e2116310118. <https://doi.org/10.1073/pnas.2116310118>
- Haydu, V. B., Camargo, J., & Bayer, H. (2015). Effects of preexperimental history on the formation of stimulus equivalence classes: A study with supporters of Brazilian soccer clubs. *Psychology & Neuroscience*, 8(3), 385–396. <https://doi.org/10.1037/h0101276>
- Hayes, S. C., Brownstein, A. J., Zettle, R. D., Rosenfarb, I., & Korn, Z. (1986). Rule-governed behavior and sensitivity to changing consequences of responding. *Journal of the Experimental Analysis of Behavior*, 45(3), 237–256. <https://doi.org/10.1901/jeab.1986.45-237>
- Hayes, S. C., Kohlenberg, B., & Hayes, L. J. (1991). The transfer of specific and general consequential functions through simple and conditional equivalence relations. *Journal of the Experimental Analysis of Behavior*, 56(1), 119–137. <https://doi.org/10.1901/jeab.1991.56-119>
- Hildum, D. C., & Brown, R. W. (1956). Verbal reinforcement and interviewer bias. *The Journal of Abnormal and Social Psychology*, 53(1), 108–111. <https://doi.org/10.1037/h0040518>
- Houmanfar, R., & Rodrigues, N. J. (2006). The metacontingency and the behavioral contingency: Points of contact and departure. *Behavior and Social Issues*, 15(1), 13–30. <https://doi.org/10.5210/bsi.v15i1.342>

- Huszár, F., Ktena, S. I., O'Brien, C., Belli, L., Schlaikjer, A., & Hardt, M. (2022). Algorithmic amplification of politics on Twitter. *Proceedings of the National Academy of Sciences*, 119(1), Article e2025334119. <https://doi.org/10.1073/pnas.2025334119>
- Insko, C. A. (1965). Verbal reinforcement of attitude. *Journal of Personality and Social Psychology*, 2(4), 621–623. <https://doi.org/10.1037/h0022485>
- Insko, C. A., & Melson, W. H. (1969). Verbal reinforcement of attitude in laboratory and nonlaboratory contexts. *Journal of Personality*, 37(1), 25–40. <https://doi.org/10.1111/j.1467-6494.1969.tb01729.x>
- Kahmann, C., & Heyer, G. (2019). Measuring context change to detect statements violating the Overton window. *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 392–396. <https://doi.org/10.5220/0008191803920396>
- Kniess, A. B., Comel, N., & Marques, F. P. J. (2024). Checagem de fatos e autoridade jornalística no Brasil: uma análise do Fato ou Fake e do Estadão Verifica. *Comunicação e Sociedade*, 46, Article e024020. [https://doi.org/10.17231/comsoc.46\(2024\).5665](https://doi.org/10.17231/comsoc.46(2024).5665)
- Lasser, J., Aroyehun, S. T., Simchon, A., Carrella, F., Garcia, D., & Lewandowsky, S. (2022). Social media sharing of low-quality news sources by political elites. *PNAS Nexus*, 1(4), Article pgac186. <https://doi.org/10.1093/pnasnexus/pgac186>
- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094–1096. <https://doi.org/10.1126/science.aao2998>
- Lelo, T. (2023). Assessing the consistency of fact-checking in political debates. *Journal of Communication*, 73(6), 587–600. <https://doi.org/10.1093/joc/jqad028>
- Levy, S., Siebold, A., Vaidya, J., Truchon, M.-M., Dettmering, J., & Mittelman, C. (2022). A look in the mirror: How the field of behavior analysis can become anti-racist. *Behavior Analysis in Practice*, 15(4), 1112–1125. <https://doi.org/10.1007/s40617-021-00630-3>
- Lindström, B., Bellander, M., Schultner, D. T., Chang, A., Tobler, P. N., & Amodio, D. M. (2021). A computational reward learning account of social media engagement. *Nature Communications*, 12(1), Article 1311. <https://doi.org/10.1038/s41467-020-19607-x>
- Linn, L. S. (1965). Verbal attitudes and overt behavior: A study of racial discrimination. *Social Forces*, 43(3), 353–364. <https://doi.org/10.2307/2574765>
- Lloyd, K. E. (1994). Do as I say, not as I do. *The Behavior Analyst*, 17(1), 131–139. <https://doi.org/10.1007/bf03392658>
- Mackinac Center for Public Policy. (2019). *The Overton window*. <https://www.mackinac.org/OvertonWindow>
- Marino-Fages, D., & Martínez-Pozo, A. (2026). *Hate in the Tropics: Political leaders and the social acceptability of online hate speech* (Documento de trabajo N.º 391). RedNIE. <https://rednie.eco.unc.edu.ar/files/DT/391.pdf>
- Massarani, L. M., Leal, T., Waltz, I., & Medeiros, A. (2021). Infodemia, desinformação e vacinas: a circulação de conteúdos em redes sociais antes e depois da COVID-19. *Liinc em Revista*, 17(1), Article e5689. <https://doi.org/10.18617/liinc.v17i1.5689>
- Matsuda, K., Garcia, Y., Catagnus, R., & Brandt, J. A. (2020). Can behavior analysis help us understand and reduce racism? A review of the current literature. *Behavior Analysis in Practice*, 13(2), 336–347. <https://doi.org/10.1007/s40617-020-00411-4>

- McGlinchey, A., & Keenan, M. (1997). Stimulus equivalence and social categorization in Northern Ireland. *Behavior and Social Issues*, 7(2), 113–128. <https://doi.org/10.5210/bsi.v7i2.310>
- Menezes, L. F. (2024, 23 de dezembro). *No segundo ano de governo, Lula desinformou mais sobre fome e educação*. Aos Fatos. <https://www.aosfatos.org/noticias/lula-2024-desinformou-fome-educacao/>
- Menezes, L. F., & Faustino, M. (2025, 7 de julho). *Como a desinformação transformou o debate político brasileiro nos últimos dez anos*. Aos Fatos. <https://www.aosfatos.org/noticias/desinformacao-transformou-debate-politico-dez-anos/>
- Merwin, R. M., & Wilson, K. G. (2005). Preliminary findings on the effects of self-referring and evaluative stimuli on stimulus equivalence class formation. *The Psychological Record*, 55(4), 561–575. <https://doi.org/10.1007/bf03395527>
- Mizael, T. M. (2018). Perspectivas analítico-comportamentais sobre a homossexualidade: análise da produção científica. *Revista Perspectivas*, 9(1), 15–28. <https://doi.org/10.18761/PAC.2017.011>
- Mizael, T. M., & de Rose, J. C. (2017). Análise do comportamento e preconceito racial: possibilidades de interpretação e desafios. *Acta Comportamentalia*, 25(3), 365–377. <https://doi.org/10.32870/ac.v25i3.61632>
- Mizael, T. M., Coelho, C. L., Rodrigues, W. C., & de Almeida, J. H. (2021a). Racial issues and behavior analysis: Experiences and contributions from Brazil. *Behavior and Social Issues*, 30(1), 495–513. <https://doi.org/10.1007/s42822-021-00071-1>
- Mizael, T. M., de Almeida, J. H., Roche, B., & de Rose, J. C. (2021b). Effectiveness of different training and testing parameters on the formation and maintenance of equivalence classes: Investigating prejudiced racial attitudes. *The Psychological Record*, 71(2), 265–277. <https://doi.org/10.1007/s40732-020-00435-w>
- Mizael, T. M., de Almeida, J. H., Silveira, C. C., & de Rose, J. C. (2016a). Changing racial bias by transfer of functions in equivalence classes. *The Psychological Record*, 66(3), 451–462. <https://doi.org/10.1007/s40732-016-0185-0>
- Mizael, T. M., Dos Santos, S. L., & De Rose, J. C. C. (2016b). Contribuições do paradigma de equivalência de estímulos para o estudo das atitudes. *Interação em Psicologia*, 20(2). <https://doi.org/10.5380/psi.v20i2.46278>
- Moreira, M. B. (Org.). (2025a). *Formação e mudança de atitudes: Pareamento ao modelo e história em quadrinhos* (1ª ed.). Instituto Walden4.
- Moreira, M. B. (Org.). (2025b). *Histórias em quadrinhos e o estudo das atitudes - Parte VII*. Instituto Walden4. https://www.walden4.com.br/livros/pdfs/moreira_2025_hq7.pdf
- Moreira, M. B., Ávila, A., Corrêa, B. F., Reis, F. G. dos, Oliveira, L. I. R. de, Pol, L. V., Silva, L. C. da, & Aguiar, S. H. A. (2023). *Histórias em quadrinhos e o estudo das atitudes - Parte II* (1ª ed.). Instituto Walden4. https://www.walden4.com.br/livros/pdfs/moreira_etal_2023_hq2.pdf
- Moreira, M. B., Cardoso, G. P., Silveira, L. A., Carvalho, L. R., & Rabelo, M. M. (2024a). *Histórias em quadrinhos e o estudo das atitudes - Parte VI*. Instituto Walden4. https://www.walden4.com.br/livros/pdfs/moreira_etal_hq6_2024.pdf
- Moreira, M. B., Lemos, I. K., Nogueira, L., Silverio, J. P. N. A., Bacelar, J. M. C., Moraes, M., Godoy, M. B. C., & da Silva, V. A. L. R. (2024b). *Histórias em quadrinhos e o estudo das atitudes - Parte III*. Instituto Walden4. https://www.walden4.com.br/livros/pdfs/moreira_etal_hq3_2024.pdf

- Moreira, M. B., Martins, A. C. M., Silva, A. F. V., Ribeiro, G. M., Faria, G. M. S., & Bezerra, J. G. N. (2024c). *Histórias em quadrinhos e o estudo das atitudes - Parte IV*. Instituto Walden4. https://www.walden4.com.br/livros/pdfs/moreira_etal_hq4_2024.pdf
- Moreira, M. B., Pova, A. M. R., Belarmino, A. J. A., Meireles, A. D. R., & Tanabe, G. Y. R. (2024d). *Histórias em quadrinhos e o estudo das atitudes - Parte V*. Instituto Walden4. https://www.walden4.com.br/livros/pdfs/moreira_etal_hq5_2024.pdf
- Moreira, M. B., Premuli, I. T., Rabelo, E. de P., Pessoa, M. B. F., & Medeiros, M. R. G. C. (2022a). *Atitudes: Histórias em quadrinhos, análises semânticas em mídias sociais e o paradigma de relações de equivalência de estímulos* (1ª ed.). Instituto Walden4. https://www.walden4.com.br/livros/pdfs/moreira_etal_2022_q0.pdf
- Moreira, M. B., Rocha, G., Lago, J., Soares, A., Santos, B. dos, Simões, C., Dias, G., Nunes, J., Alves, K., Pires, M., Jardim, M., Chadud, L., Alvarenga, O., & Souza, W. (2022b). *Preconceito: Doze experimentos e um paradigma* (1ª ed.). Instituto Walden4. https://www.walden4.com.br/livros/pdfs/moreira_etal_2022_12exp.pdf
- Moreira, M. B., Sousa, C. P., Tomasini, C., Craveiro, H. do A., Rodrigues, I. L., Dutra, I., Barros, J. F., Freire, L. R. V., D'arcanchy e Vilela, M. E., Cautela, M. E. T. C., Oliveira, P. M. D. N., Soares, S. K. S., & Andrade, V. M. (2022c). *Estudos sobre histórias em quadrinhos e atitudes - Parte I* (1ª ed.). Instituto Walden4. https://www.walden4.com.br/livros/pdfs/moreira_etal_2022_hq1.pdf
- Mosleh, M., & Rand, D. G. (2022). Measuring exposure to misinformation from political elites on Twitter. *Nature Communications*, 13(1), Article 7144. <https://doi.org/10.1038/s41467-022-34769-6>
- Mota, A. A. S., Pimentel, S. M., & Oliveira, A. V. M. G. (2023). Desordens informativas: análise de pronunciamentos de Jair Bolsonaro contra a vacinação de covid-19. *Revista Eletrônica de Comunicação, Informação & Inovação em Saúde*, 17(2), 311–331. <https://doi.org/10.29397/reciis.v17i2.3513>
- Moxon, P. D., Keenan, M., & Hine, L. (1993). Gender-role stereotyping and stimulus equivalence. *The Psychological Record*, 43, 381–394.
- Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303–330. <https://doi.org/10.1007/s11109-010-9112-2>
- Nyhan, B., Settle, J., Thorson, E., Wojcieszak, M., Barberá, P., Chen, A. Y., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Guess, A. M., Kennedy, E., Kim, Y. M., Lazer, D., Malhotra, N., Moehler, D., . . . Tucker, J. A. (2023). Like-minded sources on Facebook are prevalent but not polarizing. *Nature*, 620(7972), 137–144. <https://doi.org/10.1038/s41586-023-06297-w>
- Passarelli, D. A., Rico, A. S., & Silvestre, M. H. (2023). Contribuições da análise do comportamento para a compreensão do racismo. *Revista Brasileira de Análise do Comportamento*, 19(1). <https://doi.org/10.18542/rebac.v19i1.14941>
- Pereira, P. B. G., & Moreira, M. B. (2021a). *Atitudes em relação a presidentes do Brasil: Formação de classes de equivalência e transferência de função* (1ª ed.). Instituto Walden4. https://www.walden4.com.br/livros/pdfs/pereira_moreira_2021_presidentes.pdf
- Pereira, P. B. G., & Moreira, M. B. (2021b). Transferência de atitudes em relação a presidentes da república. *Acta Comportamentalia*, 29(4), 137–154. <https://doi.org/10.32870/ac.v29i4.80318>

- Rathje, S., Roozenbeek, J., Van Bavel, J. J., & van der Linden, S. (2023). Accuracy and social motivations shape judgements of (mis)information. *Nature Human Behaviour*, 7(6), 892–903. <https://doi.org/10.1038/s41562-023-01540-w>
- Rathje, S., Van Bavel, J. J., & van der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences*, 118(26), Article e2024292118. <https://doi.org/10.1073/pnas.2024292118>
- Recuero, R., Tavares, C., Volcan, T., Pozzebon, M., & Dutra, M. (2025). Desinformação narrada pela checagem: estudo das eleições de 2018 e 2022. *E-Compós*, 28, 1–26. <https://doi.org/10.30962/ecmps.3039>
- Regan, D. T., & Fazio, R. (1977). On the consistency between attitudes and behavior: Look to the method of attitude formation. *Journal of Experimental Social Psychology*, 13(1), 28–45. [https://doi.org/10.1016/0022-1031\(77\)90011-7](https://doi.org/10.1016/0022-1031(77)90011-7)
- Ribeiro, A. (2023, 27 de dezembro). *Em primeiro ano de mandato, Lula repete alegações enganosas sobre fome, economia e governo Bolsonaro*. Aos Fatos. <https://www.aosfatos.org/noticias/lula-alegacoes-enganosas-2023/>
- Ribeiro, A., & Menezes, L. F. (2022, 22 de dezembro). *Uma retrospectiva das mentiras do governo Bolsonaro* [Matéria em quadrinhos]. Aos Fatos. <https://www.aosfatos.org/noticias/retrospectiva-mentiras-governo-bolsonaro/>
- Ricard, J., & Medeiros, J. (2020). Using misinformation as a political weapon: COVID-19 and Bolsonaro in Brazil. *Harvard Kennedy School Misinformation Review*, 1(2). <https://doi.org/10.37016/mr-2020-013>
- Risley, T. R., & Hart, B. (1968). Developing correspondence between the nonverbal and verbal behavior of preschool children. *Journal of Applied Behavior Analysis*, 1(4), 267–281. <https://doi.org/10.1901/jaba.1968.1-267>
- Robertson, C. E., Pröllochs, N., Schwarzenegger, K., Pärnamets, P., Van Bavel, J. J., & Feuerriegel, S. (2023). Negativity drives online news consumption. *Nature Human Behaviour*, 7(5), 812–822. <https://doi.org/10.1038/s41562-023-01538-4>
- Roozenbeek, J., & van der Linden, S. (2019). Fake news game confers psychological resistance against online misinformation. *Palgrave Communications*, 5(1), Article 65. <https://doi.org/10.1057/s41599-019-0279-9>
- Roozenbeek, J., van der Linden, S., Goldberg, B., Rathje, S., & Lewandowsky, S. (2022). Psychological inoculation improves resilience against misinformation on social media. *Science Advances*, 8(34), Article eabo6254. <https://doi.org/10.1126/sciadv.abo6254>
- Rosendo, A. P., & de Melo, R. M. (2018). Transferência de função e reorganização de classes de equivalência relacionadas a gênero e profissões. *Revista Brasileira de Análise do Comportamento*, 14(1), 31–43. <https://doi.org/10.18542/rebac.v14i1.7157>
- Santos, L. F. dos. (2022, 17 de agosto). *Bolsonaro supera 6.000 declarações falsas ou distorcidas no início da campanha pela reeleição*. Aos Fatos. <https://www.aosfatos.org/noticias/bolsonaro-supera-6000-declaracoes-falsas-ou-distorcidas-no-inicio-da-campanha-pela-reeleicao/>
- Scott, W. A. (1957). Attitude change through reward of verbal behavior. *The Journal of Abnormal and Social Psychology*, 55(1), 72–75. <https://doi.org/10.1037/h0040249>
- Shea, P., Johnson, P., & Togade, D. (2023). Using relational frame theory to examine racial prejudice: A tool for educators and an appeal for future research. *Behavior Analysis in Practice*, 16(1), 102–116. <https://doi.org/10.1007/s40617-022-00767-9>

- Sidman, M., & Tailby, W. (1982). Conditional discrimination vs. Matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, 37(1), 5–22. <https://doi.org/10.1901/jeab.1982.37-5>
- Siegel, A. A., Nikitin, E., Barberá, P., Sterling, J., Pullen, B., Bonneau, R., Nagler, J., & Tucker, J. A. (2021). Trumping hate on Twitter? Online hate speech in the 2016 U.S. election campaign and its aftermath. *Quarterly Journal of Political Science*, 16(1), 71–104. <https://doi.org/10.1561/100.00019045>
- Skinner, B. F. (1957). *Verbal behavior*. Appleton-Century-Crofts.
- Spielberger, C. D., Levin, S. M., & Shepard, M. G. (1962). The effects of awareness and attitude toward the reinforcement on the operant conditioning of verbal behavior. *Journal of Personality*, 30(1), 106–121. <https://doi.org/10.1111/j.1467-6494.1962.tb02308.x>
- Street, W. R. (1994). Attitude-behavior congruity, mindfulness, and self-focused attention: A behavior-analytic reconstruction. *The Behavior Analyst*, 17(1), 145–153. <https://doi.org/10.1007/bf03392660>
- Tsipursky, G., Votta, F., & Roose, K. M. (2018). Fighting fake news and post-truth politics with behavioral science: The Pro-Truth Pledge. *Behavior and Social Issues*, 27(1), 47–70. <https://doi.org/10.5210/bsi.v27i0.9127>
- Visser, M. (1996). Voting: A behavioral analysis. *Behavior and Social Issues*, 6(1), 23–34. <https://doi.org/10.5210/bsi.v6i1.278>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>
- Weinstein, J. H., Wilson, K. G., Drake, C. E., & Kellum, K. K. (2008). A relational frame theory contribution to social categorization. *Behavior and Social Issues*, 17(1), 40–65. <https://doi.org/10.5210/bsi.v17i1.406>
- Zielinski, C. (2021). Infodemics and infodemiology: A short history, a long future. *Revista Panamericana de Salud Pública*, 45, 1. <https://doi.org/10.26633/rpsp.2021.40>

D

Verificação de fontes

NOTA DE PROCEDÊNCIA

A verificação registrada neste apêndice foi feita por uma inteligência artificial (Claude, da Anthropic), que auxiliou na redação do livro, e não pelos autores. Foi a IA que montou a planilha de afirmações, obteve ou recebeu os textos das fontes por via legal, procurou neles as frases que sustentam cada afirmação, leu o contexto e registrou os resultados e as ressalvas.

Por isso, “lida” e “não lida”, nas seções D.2 a D.4, dizem o que essa revisão conseguiu ler e conferir, e não o que os autores leram ou deixaram de ler. Uma fonte “não lida” é uma fonte que a revisão não pôde verificar: o livro não a trata como verificada e, onde o assunto exigia a fonte, o capítulo diz que ela não foi lida.

Este apêndice documenta como as afirmações dos capítulos foram conferidas contra as fontes, quantas fontes foram lidas e quais ficaram de fora, e que decisões de método afetam a leitura.

D.1 Como uma afirmação chega a “verificada”

Cada afirmação atribuída a uma fonte entra numa planilha com um identificador, a chave da fonte, o capítulo, o enunciado e as frases literais da fonte que a sustentam. Um script procura essas frases no texto integral extraído (ignorando espaços, acentos, pontuação, hifenização e ligaduras), registra a página e devolve o contexto; quem decide é a leitura do contexto, que fica registrada com a data e as ressalvas. Só então a afirmação passa a “verificada” e pode ser citada. O texto de cada capítulo marca, em comentários, as afirmações que sustentam cada parágrafo, e um segundo script confere que toda citação do texto tem referência e toda afirmação marcada existe e está verificada.

D.2 Números

O levantamento registra 168 fontes, das quais 108 foram obtidas por via legal e lidas no texto integral e 60 seguem pendentes (seção D.4). São citadas 124 obras. A planilha de afirmações tem 1207 entradas, todas registradas como verificadas, com o contexto lido. A planilha completa (referência, DOI, trecho e página, status) está em [verificacao/tabela-de-verificacao.xlsx](#).

D.3 Versões lidas que não são a do periódico

Quando a versão lida não é a publicada, ou não traz a paginação do periódico, o capítulo cita sem página e avisa. As fontes citadas no livro nessa situação:

- Ajzenman et al. (2023): versão submetida, em RePEc.
- Álvarez-Benjumea e Winter (2018): versão submetida, em MPG.PuRe.
- Bail et al. (2018): texto em XML (PMC), sem a paginação do periódico.
- Brady et al. (2021): texto em XML (PMC), sem a paginação do periódico.
- Calleja et al. (2021): texto em XML (PMC), sem a paginação do periódico.
- Ceylan et al. (2023): texto em XML (PMC), sem a paginação do periódico.
- Cinelli et al. (2021): texto em XML (PMC), sem a paginação do periódico.
- Efron e Raj (2020): manuscrito aceito, em London Business School Research Online.
- Eysenbach (2009): texto em XML (PMC), sem a paginação do periódico.
- Eysenbach (2020): texto em XML (PMC), sem a paginação do periódico.
- Ferster e Skinner (1957): edição original de 1957, Appleton-Century-Crofts, cópia da PsycBooks do acervo do autor.
- Frimer et al. (2023): manuscrito aceito, em PsyArXiv.
- Garrett e Bond (2021): texto em XML (PMC), sem a paginação do periódico.
- Guess et al. (2019): texto em XML (PMC), sem a paginação do periódico.
- Guess et al. (2020): texto em XML (PMC), sem a paginação do periódico.
- Hangartner et al. (2021): texto em XML (PMC), sem a paginação do periódico.
- Huszár et al. (2022): texto em XML (PMC), sem a paginação do periódico.
- Lasser et al. (2022): texto em XML (PMC), sem a paginação do periódico.
- Lazer et al. (2018): versão submetida, em arXiv.
- Mackinac Center (2019): página da internet, fonte interessada.

- Marino-Fages e Martínez-Pozo (2026): documento de trabalho, sem revisão de pares.
- Mota et al. (2023): versão submetida, em LA Referencia.
- Moxon et al. (1993): cópia do ResearchGate (varredura do PDF publicado).
- Nyhan e Reifler (2010): manuscrito do autor, sem a paginação da revista.
- Rathje et al. (2021): versão submetida, em Apollo.
- Roozenbeek et al. (2022): versão submetida, em Apollo.
- Shea et al. (2023): texto em XML (PMC), sem a paginação do periódico.
- Siegel et al. (2021): versão postada pelo autor.
- Skinner (1957): edição original de 1957, Appleton-Century-Crofts, cópia do acervo do autor.
- Zielinski (2021): texto em XML (PMC), sem a paginação do periódico.

D.4 Fontes não lidas

Fontes previstas no levantamento para as quais não se obteve o texto integral por via legal. Não são citadas como verificadas; onde o assunto exigiria a fonte, o capítulo diz que ela não foi lida. Das 60 fontes pendentes do levantamento, 28 têm tentativa de obtenção registrada em acesso aberto (repositório, PMC, versão do autor) em `verificacao/fontes.tsv`; as demais não tinham acesso aberto localizado, e umas e outras dependem de que o autor anexe o PDF. O último grupo da lista são livros do próprio autor, que ficam fora da exceção das regras de fontes (seção D.5).

Acesso aberto, mas o download automático foi bloqueado (24)

- Biglan, A., & Embry, D. D. (2013). A framework for intentional cultural change. *Journal of Contextual Behavioral Science*, 2(3-4), 95–104 (capítulo 12).
- Sidman, M. (2000). Equivalence relations and the reinforcement contingency. *Journal of the Experimental Analysis of Behavior*, 74(1), 127–146 (capítulo 4).
- Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A sketch of the implicit relational assessment procedure (IRAP) and the relational elaboration and coherence (REC) model. *The Psychological Record*, 60(3), 527–542 (capítulo 4).
- Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1), 13–29 (capítulos 1, 12).

- Fazio, L. K., Brashier, N. M., Payne, B. K., & Marsh, E. J. (2015). Knowledge does not protect against illusory truth. *Journal of Experimental Psychology: General*, 144(5), 993–1002 (capítulo 3).
- Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, 147(12), 1865–1880 (capítulo 3).
- Lim, C. (2018). Checking how fact-checkers check. *Research & Politics*, 5(3), Article 2053168018786848 (capítulo 2).
- Bilewicz, M., & Soral, W. (2020). Hate speech epidemic. The dynamic effects of derogatory language on intergroup relations and political radicalization. *Political Psychology*, 41(S1), 3–33 (capítulo 5).
- Ozawa, J. V. S., Woolley, S. C., Straubhaar, J., Riedl, M. J., Joseff, K., & Gursky, J. (2023). How disinformation on WhatsApp went from campaign weapon to governmental propaganda in Brazil. *Social Media + Society*, 9(1), Article 20563051231160632 (capítulo 6).
- Rossini, P., Stromer-Galley, J., Baptista, E. A., & Veiga de Oliveira, V. (2021). Dysfunctional information sharing on WhatsApp and Facebook: The role of political talk, cross-cutting exposure and social corrections. *New Media & Society*, 23(8), 2430–2451 (capítulo 6).
- Brady, W. J., Wills, J. A., Jost, J. T., Tucker, J. A., & Van Bavel, J. J. (2017). Emotion shapes the diffusion of moralized content in social networks. *Proceedings of the National Academy of Sciences*, 114(28), 7313–7318 (capítulo 8).
- Sherman, L. E., Payton, A. A., Hernandez, L. M., Greenfield, P. M., & Dapretto, M. (2016). The power of the like in adolescence: Effects of peer influence on neural and behavioral responses to social media. *Psychological Science*, 27(7), 1027–1035 (capítulo 9).
- Juul, J. L., & Ugander, J. (2021). Comparing information diffusion mechanisms by matching on cascade size. *Proceedings of the National Academy of Sciences*, 118(46), Article e2100786118 (capítulo 11).
- Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3), 554–559 (capítulo 11).
- Galhardi, C. P., Freire, N. P., Minayo, M. C. S., & Fagundes, M. C. M. (2020). Fato ou Fake? Uma análise da desinformação frente à pandemia da Covid-19 no Brasil. *Ciência & Saúde Coletiva*, 25(Suppl. 2), 4201–4210 (capítulo 11).

- Altay, S., Berriche, M., & Acerbi, A. (2023). Misinformation on misinformation: Conceptual and methodological challenges. *Social Media + Society*, 9(1), Article 20563051221150412 (capítulo 1).
- Chadwick, A., Vaccari, C., & Kaiser, J. (2025). The amplification of exaggerated and false news on social media: The roles of platform use, motivations, affect, and ideology. *American Behavioral Scientist*, 69(2), 113–130 (capítulos 8, 9).
- Bastos dos Santos, J. G., Freitas, M., Aldé, A., Santos, K., & Cardozo Cunha, V. C. (2019). WhatsApp, política mobile e desinformação: a hidra nas eleições presidenciais de 2018. *Comunicação & Sociedade*, 41(2), 307–334 (capítulo 6).
- Mercuri, K. T., & Lima-Lopes, R. E. (2020). Discurso de ódio em mídias sociais como estratégia de persuasão popular. *Trabalhos em Linguística Aplicada*, 59(2), 1216–1238 (capítulo 5).
- Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, 25(5), 388–402 (capítulos 1, 3).
- Fortuna, P., & Nunes, S. (2019). A survey on automatic detection of hate speech in text. *ACM Computing Surveys*, 51(4), 1–30 (capítulos 1, 5).
- Porter, E., & Wood, T. J. (2021). The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom. *Proceedings of the National Academy of Sciences*, 118(37), Article e2104235118 (capítulo 11).
- Holbrook, A. L., & Krosnick, J. A. (2010). Social desirability bias in voter turnout reports: Tests using the item count technique. *Public Opinion Quarterly*, 74(1), 37–67 (capítulo 7).
- Cazzamatta, R., & Santos, A. (2024). Checking verifications during the 2022 Brazilian run-off election: How fact-checking organizations exposed falsehoods and contributed to the accuracy of the public debate. *Journalism*, 25(10), 2022–2043 (capítulo 2).

Sem acesso aberto localizado (36)

- Verplanck, W. S. (1955). The control of the content of conversation: Reinforcement of statements of opinion. *The Journal of Abnormal and Social Psychology*, 51(3), 668–676 (capítulos 3, 9).
- Bem, D. J. (1965). An experimental analysis of self-persuasion. *Journal of Experimental Social Psychology*, 1(3), 199–218 (capítulo 3).
- Silver, B. D., Abramson, P. R., & Anderson, B. A. (1986). The presence of others and overreporting of voting in American national elections. *Public Opinion Quarterly*, 50(2), 228–239 (capítulos 1, 7).

- Skinner, B. F. (1945). The operational analysis of psychological terms. *Psychological Review*, 52(5), 270–277 (capítulo 3).
- Skinner, B. F. (1981). Selection by consequences. *Science*, 213(4507), 501–504 (capítulos 6, 9, 11).
- Glenn, S. S., Malott, M. E., Andery, M. A. P. A., Benvenuti, M., Houmanfar, R. A., Sandaker, I., Todorov, J. C., Tourinho, E. Z., & Vasconcelos, L. A. (2016). Toward consistent terminology in a behaviorist approach to cultural analysis. *Behavior and Social Issues*, 25(1), 11–27 (capítulos 6, 9, 11).
- Sidman, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research*, 14(1), 5–13 (capítulo 4).
- Watt, A., Keenan, M., Barnes, D., & Cairns, E. (1991). Social categorization and stimulus equivalence. *The Psychological Record*, 41(1), 33–50 (capítulo 4).
- Grey, I. M., & Barnes, D. (1996). Stimulus equivalence and attitudes. *The Psychological Record*, 46, 243 (capítulo 4).
- Dixon, M. R., Rehfeldt, R. A., Zlomke, K. R., & Robinson, A. (2006). Exploring the development and dismantling of equivalence classes involving terrorist stimuli. *The Psychological Record*, 56(1), 83–103 (capítulo 4).
- Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. *Science*, 363(6425), 374–378 (capítulo 1).
- Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188, 39–50 (capítulo 1).
- Van Bavel, J. J., Harris, E. A., Pärnamets, P., Rathje, S., Doell, K. C., & Tucker, J. A. (2021). Political psychology in the digital (mis)information age: A model of news belief and sharing. *Social Issues and Policy Review*, 15(1), 84–113 (capítulo 1).
- Hasher, L., Goldstein, D., & Toppino, T. (1977). Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior*, 16(1), 107–112 (capítulo 3).
- Uscinski, J. E., & Butler, R. W. (2013). The epistemology of fact checking. *Critical Review*, 25(2), 162–180 (capítulo 2).
- Marietta, M., Barker, D. C., & Bowser, T. (2015). Fact-checking polarized politics: Does the fact-check industry provide consistent guidance on disputed realities? *The Forum*, 13(4), 577–596 (capítulo 2).
- Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2020). Fact-checking: A meta-analysis of what works and for whom. *Political Communication*, 37(3), 350–375 (capítulos 2, 12).

- Soral, W., Bilewicz, M., & Winiewski, M. (2018). Exposure to hate speech increases prejudice through desensitization. *Aggressive Behavior*, 44(2), 136–146 (capítulo 5).
- Müller, K., & Schwarz, C. (2021). Fanning the flames of hate: Social media and hate crime. *Journal of the European Economic Association*, 19(4), 2131–2167 (capítulo 5).
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130–1132 (capítulo 6).
- González-Bailón, S., Lazer, D., Barberá, P., Zhang, M., Allcott, H., Brown, T., Crespo-Tenorio, A., Freelon, D., Gentzkow, M., Guess, A. M., Iyengar, S., Kim, Y. M., Malhotra, N., Moehler, D., Nyhan, B., Pan, J., Rivera, C. V., Settle, J., Thorson, E., . . . Tucker, J. A. (2023). Asymmetric ideological segregation in exposure to political news on Facebook. *Science*, 381(6656), 392–398 (capítulo 6).
- Gerber, A. S., Green, D. P., & Larimer, C. W. (2008). Social pressure and voter turnout: Evidence from a large-scale field experiment. *American Political Science Review*, 102(1), 33–48 (capítulo 7).
- Crockett, M. J. (2017). Moral outrage in the digital age. *Nature Human Behaviour*, 1(11), 769–771 (capítulo 8).
- Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Kennedy, E., Kim, Y. M., Lazer, D., Moehler, D., Nyhan, B., Rivera, C. V., Settle, J., Thomas, D. R., . . . Tucker, J. A. (2023). How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science*, 381(6656), 398–404 (capítulo 8).
- Eysenbach, G. (2002). Infodemiology: The epidemiology of (mis)information. *The American Journal of Medicine*, 113(9), 763–765 (capítulo 11).
- Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3), 106–131 (capítulo 12).
- Walter, N., & Murphy, S. T. (2018). How to unring the bell: A meta-analytic approach to correction of misinformation. *Communication Monographs*, 85(3), 423–441 (capítulo 12).
- Wood, T., & Porter, E. (2019). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior*, 41(1), 135–163 (capítulo 12).

- Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592(7855), 590–595 (capítulo 12).
- Lelo, T. (2023). Assessing the consistency of fact-checking in political debates. *Journal of Communication*, 73(6), 587–600 (capítulo 2).
- Tandoc, E. C., Lim, Z. W., & Ling, R. (2018). Defining “Fake News”: A typology of scholarly definitions. *Digital Journalism*, 6(2), 137–153 (capítulo 1).
- Freelon, D., & Wells, C. (2020). Disinformation as political communication. *Political Communication*, 37(2), 145–156 (capítulo 1).
- Dechêne, A., Stahl, C., Hansen, J., & Wänke, M. (2010). The truth about the truth: A meta-analytic review of the truth effect. *Personality and Social Psychology Review*, 14(2), 238–257 (capítulo 3).
- Batista Pereira, F., Bueno, N. S., Nunes, F., & Pavão, N. (2022). Fake news, fact checking, and partisanship: The resilience of rumors in the 2018 Brazilian elections. *The Journal of Politics*, 84(4), 2188–2201 (capítulos 6, 11).
- Müller, K., & Schwarz, C. (2023). From hashtag to hate crime: Twitter and antiminority sentiment. *American Economic Journal: Applied Economics*, 15(3), 270–312 (capítulo 5).
- Bernstein, R., Chadha, A., & Montjoy, R. (2001). Overreporting voting. *Public Opinion Quarterly*, 65(1), 22–44 (capítulo 7).

Livros do autor fora da exceção (3)

- Efeitos reforçadores da atenção sobre o comportamento humano (livro do autor, fora da exceção das regras de fontes) (capítulos 8, 9).
- Comportamento e práticas culturais (livro do autor, fora da exceção das regras de fontes) (capítulos 9, 10).
- Modificação do comportamento verbal através de reforçamento em intervalo variável em situação análoga à psicoterapia: uma replicação sistemática (livro do autor, fora da exceção das regras de fontes) (capítulo 9).

D.5 Decisões de método que afetam a leitura

Não se usam notícias como fonte de dado. A regra do projeto para notícias (G1, CNN Brasil, BandNews e Canal Meio, com captura de tela, URL e datas de publicação e de acesso) pede coleta de páginas e imagens que não se pôde fazer de modo auditável; os casos de fala são os documentados em artigos e relatórios cujo texto integral foi lido, citados com a página. A regra continua valendo para quem acrescentar notícias.

As falas do Capítulo 3 são casos ilustrativos, e não amostra. A amostra comparável do Capítulo 2 (seção 2.9) tem regra fixada e não foi coletada: a Lupa e o Aos Fatos proíbem rastreadores de inteligência artificial em seus robots.txt, o feed público de ClaimReview cobre a Lupa de forma muito desigual e o relatório da Lupa não discrimina os ciclos de Lula. Antes de qualquer requisição a um site, o robots.txt foi lido num passo próprio.

Duas obras de livro entram por exceção, aprovada pelo autor: Skinner (1957, *Verbal behavior*) e Ferster e Skinner (1957, *Schedules of reinforcement*), lidos em cópias do acervo, com as páginas impressas. Os demais livros ficam fora, salvo os do autor sobre atitudes e sobre histórias em quadrinhos. Três outros livros do autor, sobre os efeitos reforçadores da atenção, sobre práticas culturais e uma replicação sistemática de modificação do comportamento verbal, não estão nessa exceção: não foram lidos e não são citados, e reforçariam os Capítulos 8 a 10 se o autor ampliar a exceção. O manual “O conceito de atitude na Análise do Comportamento”, do autor, segue inédito e não é citado; a definição de atitude em camadas (Capítulo 4, seção 4.1) entra ancorada em obras publicadas.

Os recortes das imagens do relatório da Lupa de que o Capítulo 3 lê as classificações estão em verificacao/evidencias/, fora do texto do livro.

D.6 Deslizes de procedimento registrados

Dois deslizes de método são registrados porque se afirma que o robots.txt foi lido antes de qualquer requisição. Em 11/10/2026, para obter metadados de Frimer et al. (2023), uma requisição foi enviada a api.osf.io antes de conferir o robots.txt, que lá é Disallow: /; voltaram só metadados do resumo, que não foram usados, e o texto lido é o do PDF obtido em osf.io/download. Em 14/10/2026, ao tentar obter o texto de duas fontes pelo PMC, uma requisição ao efetch de eutils.ncbi.nlm.nih.gov foi feita no mesmo comando que imprimia o robots.txt, também Disallow: /; voltaram só metadados sem corpo, nada foi usado e os arquivos foram apagados. Desde então o robots.txt é lido num passo próprio, e as E-utilities do NCBI não são usadas.

EDITORA

Editora do Instituto Walden4

A Editora do Instituto Walden4 se dedica à promoção e difusão de conhecimentos avançados na esfera da Análise do Comportamento, assim como em outras áreas práticas baseadas em evidências, abrangendo tanto aspectos científicos quanto profissionais. Comprometida com a democratização do acesso à informação, oferece uma ampla gama de livros de forma gratuita, e todos os títulos estão acessíveis em formato digital.

CONSELHO EDITORIAL

Dr. Gleidson Gabriel da Cruz

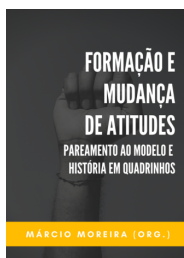
Dr. Márcio Borges Moreira

CONTATO

secretaria@walden4.com.br

walden4.com.br • [@instituto.walden4](https://www.instagram.com/instituto.walden4)

Outros livros da editora do Instituto Walden4



Formação e Mudança de Atitudes: Pareamento ao Modelo e História em Quadrinhos

Márcio Borges Moreira (org.) · 2025



Preconceito: doze experimentos e um paradigma

Márcio Borges Moreira · 2021



Atitudes: histórias em quadrinhos, análises semânticas em mídias sociais e o paradigma de relações de equivalência de estímulos

Márcio Borges Moreira · 2021



Atitudes em relação a presidentes do Brasil: formação de classes de equivalência e transferência de função

Pedro Bufara Gonzaga Pereira, Márcio Borges Moreira · 2021



Efeitos reforçadores da atenção sobre o comportamento humano

Márcio Borges Moreira · 2021



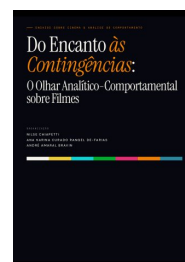
Comportamento e Práticas Culturais

2015



Inteligência Artificial na Prática Clínica

Medeiros et al. · 2026



Do Encanto às Contingências

2026



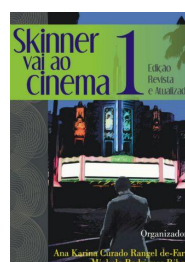
Uma história de aprendizagem operante

Márcio Borges Moreira, Lucas Couto de Carvalho · 2017



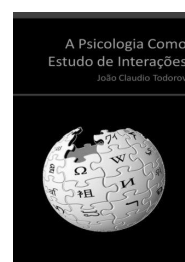
Experimentos clássicos em Análise do Comportamento

2016



Skinner Vai ao Cinema (Vol. 1)

2015



A Psicologia como Estudo de Interações

João Claudio Todorov · 2015

Conheça o catálogo completo, com 65 títulos, muitos gratuitos

walden4.com.br/livros

