ANÁLISIS DE LA CONDUCTA

Un análisis operante de

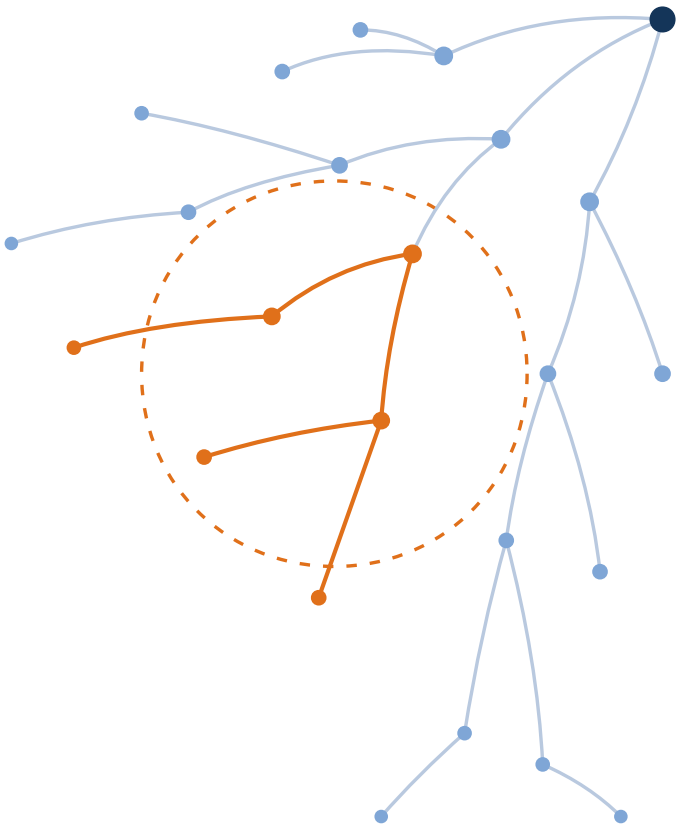
los discursos de odio y la desinformación



Márcio Moreira · Carlos Augusto de Medeiros · Cristiano Coelho
Diogo Conque Seco-Ferreira · Frederico Neves Condé · Lucas Ferraz Córdova

iW4 instituto WALDEN4
walden4.com.br

instituto WALDEN4
walden4.com.br



ANÁLISIS DE LA CONDUCTA

Un análisis operante de

los discursos de odio y la desinformación

Márcio Moreira
Carlos Augusto de Medeiros
Cristiano Coelho
Diogo Conque Seco-Ferreira
Frederico Neves Condé
Lucas Ferraz Córdova

Instituto Walden4
Brasília, DF · 2026

FICHA CATALOGRÁFICA

CDD-150

Datos Internacionales de Catalogación en la Publicación
(Elaborados por la editorial)

Un análisis operante de los discursos de odio y la desinformación [libro electrónico] / Márcio Borges Moreira ... [et al.] ; traducción del portugués de Brasil de Márcio Borges Moreira. -- Brasília, DF : Instituto Walden4, 2026.

PDF

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

Traducción de: Uma análise operante de discursos de ódio e desinformação (ISBN 978-85-65721-74-5).

Bibliografía.

ISBN 978-85-65721-87-5

1. Conducta agresiva 2. Conducta - Análisis 3. Conducta - Aspectos psicológicos 4. Comunicación - Aspectos políticos 5. Comunicación - Aspectos sociales 6. Desinformación 7. Fake news 8. Odio - Aspectos sociales 9. Política - Aspectos sociales 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. Conducta : Análisis : Psicología 150

Ficha catalográfica elaborada por la editorial

ISBN 978-85-65721-87-5
9 788565 721875

Instituto Walden4
Brasília, DF · 2026

La idea de este libro

La idea de hacer este libro surgió de una publicación en redes sociales que Rafael Nunes Félix envió a Carlos Augusto de Medeiros. Dejamos constancia aquí del crédito de Rafael por la idea, con nuestro agradecimiento.

Rafael Nunes Félix es psicólogo egresado del CEUB, con interés en la investigación y la práctica clínica en Análisis de la Conducta. Fue docente en la Universidade de Brasília (UnB), donde impartió clases de Psicología de la Personalidad desde la perspectiva del Análisis de la Conducta, y hoy es estudiante no regular del Programa de Posgrado en Ciencias del Comportamiento de la misma universidad. Investiga el descuento por demora y la resurgencia conductual, con foco en los procesos básicos de la conducta y en sus aplicaciones.

Sobre los autores

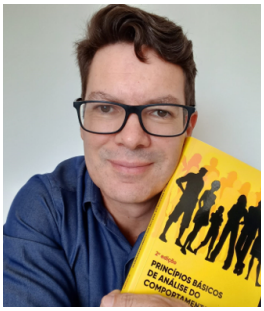
Márcio Moreira



Márcio Borges Moreira es doctor en Ciencias del Comportamiento por la Universidade de Brasília, y magíster en Psicología y psicólogo por la Pontifícia Universidade Católica de Goiás. Es profesor del grado y de la Maestría en Psicología del Centro Universitário de Brasília (UniCEUB), donde supervisa prácticas clínicas en el Centro de Formação de Psicólogos (CENFOR). Es también director del Instituto Walden4, clínica especializada en servicios basados en ABA para niños y adolescentes con trastorno del espectro autista (TEA).

Coautor del libro *Princípios Básicos de Análise do Comportamento*, tiene capítulos y artículos científicos publicados en revistas brasileñas e internacionales, con foco en Análisis Aplicado de la Conducta, conducta verbal e intervenciones basadas en evidencia. Sus proyectos de investigación actuales estudian la automatización de entrenamientos para profesionales y paraprofesionales que trabajan con personas diagnosticadas con autismo, el uso de tecnología e inteligencia artificial en la enseñanza de conceptos psicológicos, la formación y el cambio de actitudes prejuiciosas y la medición de conductas en la prestación de servicios de psicología.

Carlos Augusto de Medeiros



El Dr. Carlos Augusto de Medeiros concluyó su doctorado con honores en el programa de Ciencias del Comportamiento de la Universidade de Brasília - UnB en 2003. Obtuvo la maestría en Psicología también en la UnB en 1999, tras concluir el grado en Psicología en 1997 en la misma institución. Fue profesor asistente y adjunto en la Universidade Federal de Minas Gerais - UFMG de 2000 a 2003. En el Centro Universitário de Brasília - UniCEUB, de 2003 a 2024, ejerció las funciones de profesor de grado, supervisor de prácticas profesionales, coordinador, profesor y director de tesis del programa de Maestría en Psicología. Impartió diversos módulos en cursos de posgrado en Análisis de la Conducta en Brasil, y fue supervisor de prácticas clínicas en el Instituto Brasiliense de Análise do Comportamento - IBAC. A lo largo de su ex-

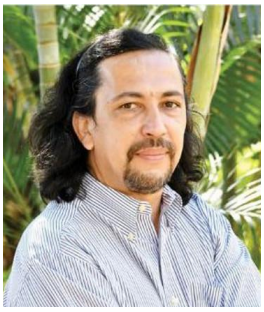
tensa carrera, impartió asignaturas relacionadas con el Análisis de la Conducta, Introducción a la Psicología y Clínica Analítico-Conductual, dirigió proyectos de investigación y supervisó prácticas de investigación y profesionales.

Desde 1997 trabaja como psicólogo clínico, con énfasis en terapia individual de adultos y terapia de pareja. Actualmente es uno de los directores de Pragmática: Psicoterapia e Cursos, una clínica de psicología que ofrece cursos regulares de extensión en Psicología Clínica.

Desarrolló un modelo de terapia analítico-conductual denominado Psicoterapia Conductual Pragmática, un modelo basado en la filosofía conductista radical y comprometido con el desarrollo de la independencia y la autonomía de los clientes.

A lo largo de su larga carrera, el Dr. Carlos Augusto de Medeiros se ha dedicado a la formación de analistas de la conducta en Brasil. Es coautor del best seller "Princípios Básicos de Análise do Comportamento". Tiene 22 artículos, dos libros y 43 capítulos de libros publicados. Sus publicaciones se relacionan con la conducta verbal, las relaciones de equivalencia, la conducta gobernada por reglas, la clínica conductual y el Análisis de la Conducta propiamente dicho.

Cristiano Coelho



Es graduado, magíster y doctor en Psicología por la Universidade de Brasília. Fue Profesor Sustituto en la UnB (1999 a 2000), Profesor Colaborador del Instituto de Educação Superior de Brasília - IESB (2009 a 2010), Supervisor de Encuestas del IBGE, el instituto nacional de estadística de Brasil, en el estado de Goiás (2010 a 2014), Profesor Colaborador del MBA en Economía Conductual de la ESPM y miembro de la junta directiva 2019-2020 de la Associação Brasileira de Psicologia e Medicina Comportamental (ABPMC). Actualmente es profesor Adjunto I de la Pontifícia Universidade Católica de Goiás, Coordinador del Programa de Posgrado Stricto Sensu en Psicología de la PUC-Goiás (Maestría y Doctorado) y terapeuta conductual. Ha realizado investigaciones y dirigido a estudiantes de iniciación científica, Maestría y Doctorado principalmente en los siguientes temas: conducta de elección, autocontrol e impulsividad, habilidades sociales y análisis de la conducta del consumidor.

Diogo Conque Seco-Ferreira



Diogo Conque Seco-Ferreira es psicólogo por la Universidade de Brasília (UnB), donde también obtuvo la maestría en Psicología y el doctorado en Ciencias del Comportamiento, con una investigación sobre la conducta del consumidor, y realizó un posdoctorado en economía conductual en la Universidad de Tilburg, en los Países Bajos. Es profesor asociado del Departamento de Psicología de la Universidade Federal de Sergipe (UFS).

Trabaja en la interfaz entre el Análisis de la Conducta, la psicología económica y la economía conductual, con investigaciones sobre la conducta del consumidor, el análisis conductual de fenómenos sociales y las ciencias conductuales aplicadas a las políticas públicas. Es coautor de "Economia comportamental: uma introdução para analistas do comportamento", publicado en la Revista Interamericana de Psicología.

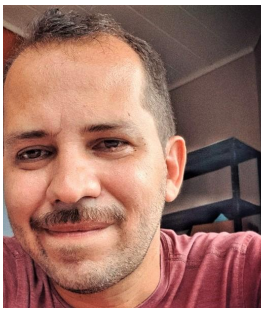
Frederico Neves Condé



Doctor, magíster y graduado en Psicología Social, del Trabajo y de las Organizaciones por la Universidade de Brasília (UnB). Es investigador-tecnólogo en Información y Evaluaciones Educativas del Inep (el instituto nacional de estudios e investigaciones educativas de Brasil), donde dirigió el Servicio de Cálculo de Indicadores de Instrumentos Contextuales (2022–2023), fue Director de Evaluación de la Educación Básica (2014–2015) y fue uno de los desarrolladores del Inse, el indicador de nivel socioeconómico del Saeb (2019 y 2021). Fue asesor técnico de la Presidencia Ejecutiva y de la Defensoría del Usuario de Anatel (la agencia reguladora de las telecomunicaciones en Brasil), analista de información social en el MDS (Ministerio de Desarrollo Social), consultor del PNUD y de la Unesco y consejero del CRP-01 (Consejo Regional de Psicología del Distrito Federal). Fue docente de grado y de posgrado (IESB y UniCEUB) y director-presidente de Psicométrica Consultoria e Pesquisa. Sus áreas de actuación son la psicometría y la evaluación psicológica, educativa y social, y se dedica a la Teoría de Respuesta al Ítem, a la elaboración de instrumentos y al modelado de indicadores.

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

Lucas Ferraz Córdova



Lucas Ferraz Córdova es psicólogo por la Universidade Federal de Minas Gerais (UFMG), magíster en Psicología y doctor en Ciencias del Comportamiento por la Universidade de Brasília (UnB), con una tesis sobre el efecto del entrenamiento sucesivo en la conducta de transposición entre los operantes verbales mando y tacto. Es profesor de la Universidade Federal dos Vales do Jequitinhonha e Mucuri (UFVJM).

Trabaja en Análisis de la Conducta, con énfasis en conducta verbal, filosofía del Conductismo Radical y relaciones entre lenguaje, contexto y explicación científica. Coordina proyectos sobre los determinantes ambientales e históricos de la conducta de explicar, que investigan las variables que afectan el establecimiento y el mantenimiento del repertorio verbal, además de estudios experimentales sobre entrenamiento de operantes verbales, historia de reforzamiento y efectos de las audiencias sobre el discurso explicativo.

Índice

Presentación 13

Alcance y límites 13
Las reglas de lectura 13
Cómo leer los recuadros y los términos 14
El mapa del libro 15
Orden de lectura 16
Referencias *16*

Nota sobre el proceso de producción 17

Lo que vino de los autores 17
Lo que hizo la IA 18
Ejemplos de prompts 20
Lo que hicimos después 25
Límites del proceso 26
Una lectura conductual del propio proceso 26
Responsabilidad 27
La edición en español 27

PARTE I **El fenómeno y los datos** 28

Capítulo 1 — Desinformación y discurso de odio: definiciones, medidas y datos 29

1.1 Definiciones, y lo que cada una exige medir 29
1.2 Prevalencia: cuánto circula y quién comparte 31
1.3 Las élites políticas como fuente 33
1.4 Posicionamiento político y propensión a la información falsa: lo que muestran los estudios correlacionales 34
1.5 Limitaciones: lo que los diseños no permiten 37
1.6 Lo que se sabe sobre Brasil 39
1.7 Discurso de odio: definiciones, medidas y datos 40
1.8 Interpretaciones rivales y límites 42
Síntesis *43*
Referencias *43*

Capítulo 2 — Lo que dicen las agencias de verificación (y lo que no pueden decir) 47

2.1 Dos series de contingencias 48
2.2 Lo que cuenta cada fuente 50
2.3 Por qué la proporción entre las verificadas no es la proporción entre las dichas 51
2.4 Lo que Lupa permite decir 53
2.5 Lo que Aos Fatos permite decir 55
2.6 Dichos y menciones: de quién es la falsedad 57
2.7 ¿Concuerdan las agencias entre sí? 57
2.8 Lo que la verificación no alcanza 58
2.9 Protocolo para una muestra comparable 59

Síntesis *61*
Referencias *61*

PARTE II **Análisis operante de los discursos** 63

Capítulo 3 — Declaraciones falsas como operantes verbales **64**
3.1 El discurso del político como conducta verbal 65
3.2 La ficha de análisis 66
3.3 ¿Tacto o intraverbal? El continuo y la forma de tacto 69
3.4 Tacto distorsionado: cuando la consecuencia del oyente desplaza el control por el evento 72
3.5 Repetición: la fuerza de la afirmación repetida 75
3.6 Autoclíticos de fuente y de certeza: lo que se añade y lo que se quita 77
3.7 La declaración como regla para la audiencia 81
3.8 Casos: las mismas cuatro frases, con el mismo método 82
3.9 Interpretaciones rivales y cómo decidir entre ellas 86
Síntesis *87*
Referencias *88*

Capítulo 4 — Actitudes políticas y prejuicio como clases de equivalencia **91**
4.1 La actitud como conducta verbal evaluativa 92
4.2 Cómo una palabra adquiere función sin contacto con el referente 93
4.3 Transferencia de función 97
4.4 El prejuicio en los experimentos 103
4.5 Actitudes hacia presidentes de Brasil 107
4.6 "Nosotros" y "ellos": relaciones de oposición y de comparación 121
4.7 Cambio de actitud: reorganización de las clases y cómics 123
4.8 Medidas, interpretaciones rivales y límites 133
Síntesis *136*
Referencias *137*

Capítulo 5 — Discurso de odio: mando disfrazado y control aversivo **142**
5.1 Lo que produce la conducta del hablante: aprobación, retirada de amenazas y funciones que varían con el contexto 143
5.2 Mando disfrazado: la opinión que funciona como instrucción 145
5.3 Control aversivo dirigido a un grupo 148
5.4 El rótulo del grupo y la transferencia de función aversiva 150
5.5 Del discurso al acto: lo que los datos permiten afirmar 151
5.6 Un caso con clasificación externa: la declaración de 2011 y lo que vino después 156
5.7 Interpretaciones rivales y límites 158
Síntesis *159*
Referencias *160*

Capítulo 6 — Burbujas digitales como comunidades verbales y la evitación del contracontrol **163**
6.1 La burbuja como comunidad verbal: audiencia y fuente de reforzamiento 164
6.2 Lo que dicen los datos sobre las cámaras de eco (y dónde divergen) 166
6.3 "Realidades virtuales": cuando el control verbal sustituye al contacto con el evento. 170
6.4 Contracontrol y evitación del contracontrol 171
6.5 El caso brasileño: WhatsApp, elecciones y pandemia 176

6.6 Quién da “me gusta”, quién responde: un análisis exploratorio de dos videos políticos en Instagram ... 179
6.7 Análisis semánticos en redes sociales ... 194
6.8 Interpretaciones rivales y límites ... 195
Síntesis ... *198*
Referencias ... *198*

Capítulo 7 — El voto como conducta verbal ... 201
7.1 El voto según la definición de conducta verbal ... 202
7.2 ¿Mando, tacto o intraverbal? Funciones posibles del voto ... 203
7.3 El voto secreto como retirada de la audiencia ... 205
7.4 Declarar y votar: actitudes verbales y actos ... 207
7.5 Consecuencias sociales de votar: los experimentos de campo ... 212
7.6 Nombres y números de candidatos en clases de equivalencia ... 215
7.7 Interpretaciones rivales y límites ... 217
Síntesis ... *218*
Referencias ... *219*

PARTE III La ecología de la información 221

Capítulo 8 — Economía de la atención: el algoritmo como programador de contingencias ... 222
8.1 La atención como reforzador ... 223
8.2 Lo que muestran los datos sobre el contenido emocional, moralizado y negativo ... 225
8.3 El algoritmo como programador: cuánta audiencia recibe cada emisión ... 228
8.4 Reforzamiento intermitente e interacción ... 232
8.5 Experimentos con el propio algoritmo: lo que cambió y lo que no cambió ... 237
8.6 Interpretaciones rivales y límites ... 240
Síntesis ... *241*
Referencias ... *242*

Capítulo 9 — El ciclo productor-audiencia ... 245
9.1 Las opiniones pueden reforzarse: los experimentos de 1956 a 1969 ... 246
9.2 Lo que recibe el productor: “me gusta”, veces compartido y alcance ... 252
9.3 Lo que hace la audiencia y lo que la mantiene ... 255
9.4 El ciclo: cada lado moldea al otro ... 257
9.5 Indignación moral aprendida por reforzamiento: la evidencia ... 258
9.6 El político como productor de contenido ... 261
9.7 La metacontingencia del ecosistema: productor, audiencia, plataforma y anunciante ... 262
9.8 Interpretaciones rivales y límites ... 264
Síntesis ... *266*
Referencias ... *266*

Capítulo 10 — La ventana de Overton: lo que se puede decir sin ser castigado ... 269
10.1 El concepto, el origen y la traducción propuesta ... 270
10.2 De la ventana de la política a la ventana del enunciado: lo que desplaza el borde ... 273
10.3 Información sobre lo que los otros toleran: el experimento de Pittsburgh ... 274
10.4 Dos elecciones, dos medidas: Brasil, 2018, y Estados Unidos, 2016 ... 277

10.5 Políticos y redes sociales: quién habla desde arriba y qué recibe 281
10.6 Sanción, contracontrol y lo que mantiene el borde en su lugar 284
10.7 Cómo situar un dicho en relación con la ventana 287
10.8 Interpretaciones rivales y límites 288
Síntesis *290*
Referencias *291*

Capítulo 11 — Infodemiología: cómo se propaga la información y qué se mide 293

11.1 Infodemiología e infovigilancia 294
11.2 Medidas de difusión: alcance, exposición, reacción y la estructura de la cascada 296
11.3 ¿Lo falso se propaga más rápido que lo verdadero? 299
11.4 Modelos epidemiológicos aplicados a la información 302
11.5 La infodemia de covid-19, en el mundo y en Brasil 305
11.6 Lo que la infodemiología propone hacer 308
11.7 Preguntas para leer un estudio de difusión 310
Síntesis *310*
Referencias *311*

PARTE IV **Intervenciones** 313

Capítulo 12 — Intervenciones: qué se puede cambiar, en cuatro niveles, y con qué evidencia 314

12.1 Corregir después: la corrección, la advertencia y el "efecto rebote" 317
12.2 Corregir antes: la inoculación como entrenamiento en discriminación 320
12.3 Estímulos que señalan precisión 324
12.4 Ante quien habla: contradiscurso, compromiso público y autoclíticos de fuente 327
12.5 Reducir el prejuicio mediante equivalencia y mediante cómics 331
12.6 Cambiar las contingencias de la plataforma 334
12.7 Diseño cultural y los límites éticos de quien diseña 339
Síntesis *342*
Referencias *343*

Apéndice A — Ficha de análisis del discurso 346

Apéndice B — Conceptos en uso 348

Apéndice C — Referencias 352

Apéndice D — Verificación de fuentes 361

D.1 Cómo una afirmación llega a "verificada" 361
D.2 Números 362
D.3 Versiones leídas que no son la de la revista 362
D.4 Fuentes no leídas 363
D.5 Decisiones de método que afectan la lectura 369
D.6 Deslices de procedimiento registrados 369

Presentación

Este libro reúne una pregunta y un método. La pregunta es qué mantiene, en términos de contingencias, lo que políticos, plataformas y audiencias dicen y reenvían cuando el enunciado es falso o alguien lo clasifica como discurso de odio. El método es el del Análisis de la Conducta aplicado a la conducta verbal: en lugar de inferir lo que el hablante tenía en mente, se pregunta qué variables controlan la emisión y qué produce esta en quien escucha. Skinner (1957) da el vocabulario; la literatura experimental sobre actitudes, equivalencia y contingencias culturales da lo que se sabe y lo que todavía no se sabe.

Fue escrito para profesionales que ya conocen el Análisis de la Conducta: analistas de la conducta, psicólogos, investigadores. Por eso no hay una parte de fundamentos. Cada concepto entra, en un recuadro "Concepto en uso", en el punto en que el análisis pasa a necesitarlo, y el Apéndice B los reúne con el capítulo en que se presentó cada uno.

Alcance y límites

Aquí no se clasifican declaraciones como falsas, no se señala quién miente y no se defiende un campo político. Quien clasifica es una agencia de verificación, una decisión judicial, una plataforma o un estudio, y el texto dice cuál. Tampoco se puso a prueba ninguna hipótesis: se leyeron estudios de terceros y, donde se propone una lectura operante, se dice que es una lectura. Lo que se obtiene de aquí es un modo de leer el dato y de formular la pregunta que falta, y no una respuesta sobre quién es más deshonesto.

Dos consecuencias aparecen en todos los capítulos. La primera es que la clasificación de un operante verbal (mando, tacto, intraverbal, tacto distorsionado) es siempre interpretación, y nunca dato: el texto dice qué variables de control la sostienen y cuál sería la interpretación rival. La segunda es que la cobertura es desigual. Buena parte de los estudios son de lengua inglesa y de Estados Unidos, provienen sobre todo de Twitter y de la covid-19, y sobre Brasil hay pocos; cada capítulo dice lo que eso permite y lo que no permite concluir.

Las reglas de lectura

Cinco reglas valen de principio a fin.

La primera es el mismo método para todos los campos políticos. Donde se analiza lo que dijo un político, se busca el caso espejo del otro campo con el mismo criterio, y cuando no se encuentra, se dice que no lo hay. La regla de muestra comparable del Capítulo 2 (sección 2.9) se fijó antes de cualquier recolección, para que la elección de los casos no dependiera del resultado; la recolección no se hizo, y el capítulo dice por qué.

La segunda es que la falsedad es siempre de la agencia que la clasificó. "Falso" aparece con el nombre de la agencia, la etiqueta que usó y la fecha, y nunca como una propiedad que quien escribe atribuye a la frase.

La tercera es que "discurso de odio" es una etiqueta externa. Solo se llama así al enunciado clasificado por una instancia externa, con la instancia nombrada. Sin ella, se describe la función del enunciado, y no se lo etiqueta.

La cuarta es la separación entre dato, interpretación e hipótesis. Todo análisis de datos o de discurso trae un recuadro con las tres partes: lo que muestra la fuente, lo que se lee en ello y lo que queda como hipótesis, con lo que haría falta para ponerla a prueba.

La quinta trata de las fuentes. Solo entra como fuente lo que se leyó en el texto completo, y cada afirmación atribuida a una fuente se localizó en ella y se leyó en contexto; las planillas de ese trabajo se describen en el Apéndice D. Donde la fuente no pudo obtenerse por vía legal, se escribe que no fue leída, y no se la cita como verificada. Las páginas citadas son las de la revista cuando el texto leído las trae; cuando no las trae, la cita va sin página, y el capítulo advierte qué versión se leyó (manuscrito aceptado, documento de trabajo, texto en XML).

Cómo leer los recuadros y los términos

Hay tres tipos de recuadro. El recuadro "Dato · Interpretación · Hipótesis" cierra los análisis. El recuadro "Concepto en uso" presenta un concepto de la literatura, con la fuente y el punto en que pasa a usarse; cuando un capítulo lo retoma, el recuadro dice "retomada" y remite al capítulo de origen. El recuadro "Definición propuesta" marca el término que se propone aquí y que ninguna fuente leída usa, como la ventana de una audiencia (Capítulo 10) y la distinción entre alcance, exposición y reacción (Capítulo 11); "evitación del contracontrol" (Capítulo 6) es una propuesta hecha aquí, a partir de Delprato (2002), y viene siempre como hipótesis.

Las citas siguen el formato APA: "Autor (año, p. X)". Las referencias de cada capítulo están consolidadas en el Apéndice C. En el texto, "fue leído" y "no fue leído" quieren decir exactamente esto: el texto completo fue o no fue leído por quienes escribieron este libro.

Los estudios citados traen estadísticas, y el texto las explica la primera vez que las usa; este párrafo es para tenerlo a mano. El valor de *p* es la probabilidad de obtener una diferencia tan grande como la observada, o mayor, si en realidad no hubiera diferencia; por convención, un *p* por debajo de 0,05 hace que la diferencia se llame "significativa", lo que informa que difícilmente se debe al azar y no que sea grande o importante. *t* y *F* son los estadísticos de las pruebas que producen el *p*. La *d* y la *h* de Cohen miden el tamaño de un efecto, la diferencia entre medias en el primer caso y entre proporciones en el segundo, y los valores cercanos a 0,2 son pequeños. La razón de posibilidades (*odds ratio*, OR) compara la posibilidad de un desenlace en dos grupos: 1 significa posibilidades iguales. La desviación estándar es la variación típica de una medida, y el intervalo de confianza del 95% es el rango de valores plausibles para la estimación. El R^2 es la fracción de la variación de los datos que explica un modelo, y la exactitud, la precisión, la exhaustividad y el *F*1 describen los aciertos de un clasificador de texto. Cuando un estudio no tiene grupo de control, o mide solo lo que las personas declaran, el texto lo dice.

El mapa del libro

El libro se divide en cuatro partes y cuatro apéndices. La Parte I, "El fenómeno y los datos", tiene dos capítulos. El Capítulo 1 reúne las definiciones y los estudios de prevalencia y correlacionales, con las limitaciones de cada medida, y fija la regla de que lo que se mide es función del instrumento. El Capítulo 2 trata de las agencias de verificación: lo que mide y lo que no mide la proporción de falsas publicada por una agencia, y por qué no se la compara con la de otra sin las salvedades del capítulo.

La Parte II, "Análisis operante de los discursos", tiene cinco. El Capítulo 3 es el núcleo: presenta la ficha de análisis y los operantes verbales que usan los siguientes, y la aplica a cuatro declaraciones verificadas. El Capítulo 4 trata de las actitudes y del prejuicio como clases de equivalencia. El 5 trata del discurso de odio como mando disfrazado y control aversivo; el 6, de las burbujas digitales como comunidades verbales y de la evitación del contracontrol; el 7, del voto como conducta verbal.

La Parte III, "La ecología de la información", tiene cuatro. El Capítulo 8 trata de la economía de la atención y del algoritmo como programador de contingencias; el 9, del ciclo productor–audiencia; el 10, de la ventana de Overton,

definida por lo que una audiencia castiga; el 11, de la infodemiología y de las medidas de difusión. La Parte IV tiene un capítulo, el 12, sobre las intervenciones que se derivan del análisis, en cuatro niveles, y la evidencia de cada una. Los apéndices traen la ficha de análisis (A), los conceptos en uso (B), las referencias consolidadas (C) y la verificación de fuentes, con las lagunas (D).

Orden de lectura

Los Capítulos 1 y 2 pueden leerse por separado, por quien quiera solo los datos y la lectura crítica de las verificaciones. El Capítulo 3 es el punto de apoyo de los demás: quien vaya a leer el análisis de declaraciones y de discurso de odio, en los Capítulos 5 y 7, debe pasar por él y por el 4. Fuera de eso, los capítulos son autónomos, con remisiones explícitas a lo que usan de otro. El Capítulo 12 se lee mejor al final, porque usa el marco del entrenamiento en discriminación, las distinciones de los Capítulos 8 y 9 y la definición de alcance, exposición y reacción del Capítulo 11.

Quien use el libro como herramienta de análisis encuentra preguntas listas en cuatro puntos: la ficha del Capítulo 3 (sección 3.2), reproducida en el Apéndice A, para leer una declaración; las del Capítulo 10 (sección 10.7), para situar una declaración en una ventana; las del Capítulo 11 (sección 11.7), para leer una medida de difusión; y las del Capítulo 12 (sección 12.7), para leer una intervención. La ficha es la de mayor uso: transforma lo que describen los capítulos en una secuencia de preguntas sobre la declaración, la audiencia y las consecuencias, y termina siempre en la que importa, que es qué mantiene la declaración.

Referencias

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 el proceso de producción

Este libro se escribió con un uso intensivo de inteligencia artificial generativa, y esta nota describe cómo. Los autores somos nosotros, los seis firmantes (Márcio Moreira, Carlos Augusto de Medeiros, Cristiano Coelho, Diogo Conque Seco-Ferreira, Frederico Neves Condé y Lucas Ferraz Córdova): definimos el problema, el marco interpretativo y los criterios de análisis, y respondemos por cada afirmación publicada. La IA redactó las versiones iniciales de todos los capítulos, de la Presentación y de los Apéndices A y B, y condujo la verificación de las fuentes; nosotros las revisamos, modificamos, ajustamos y comprobamos, entre el 19 y el 26 de octubre de 2026. La conducción de la IA, desde el pedido original hasta el último pedido de revisión, la hizo toda Márcio Moreira.

Preferimos describir el proceso en detalle a resumirlo en una línea de agradecimiento. El lector tiene derecho a saber cómo se produjo el texto para evaluarlo, y una descripción vaga serviría mal tanto a la transparencia como al trabajo que de hecho hicimos. Por el mismo motivo, no llamamos a este un libro "generado por IA": la expresión describe mal un proceso en el que el problema, el método y el juicio final fueron humanos de principio a fin.

Lo que vino de los autores

Lo que orienta el libro antecede a la redacción con IA y vino de nosotros, en el pedido original del 10 de octubre de 2026 y en los pedidos siguientes.

- **El problema.** La elección del discurso de odio y de la desinformación como objeto, con foco en las declaraciones de políticos brasileños, y la pregunta central: qué puede decir un análisis skinneriano sobre este fenómeno a partir de los datos que la ciberpsicología y las áreas vecinas (ciencia política, comunicación, infodemiología) ya han producido.
- **El marco interpretativo.** Una posición skinneriana explícita, fundada en la conducta verbal (*Verbal behavior*, de Skinner, y *Schedules of reinforcement*, de Ferster y Skinner, que abrimos como excepción a la regla de que solo se citan artículos), combinada con el análisis contextual de Bernard Guerin. Del artículo de Guerin que indicamos como base entran las actitudes y las creencias como conducta verbal mantenida por comunidades verbales, las funciones de tacto, intraverbal y mando de un mismo enunciado y el papel de los autoclíticos en la negociación social de lo que se

dice; de otro trabajo suyo, el análisis del conocimiento como construcción social y la condición en la que un grupo pequeño y cohesionado sostiene lo que dice. A esto se suman la equivalencia de estímulos y la transferencia de función, trabajadas en libros nuestros sobre actitudes y sobre cómics (historietas). Este marco llegó a las herramientas por el pedido original, por el artículo base, por esos libros y por las copias de Skinner que proporcionamos.

- **La selección de la literatura.** Fijamos las reglas: solo artículos científicos, más nuestros libros sobre actitudes y cómics y las dos obras de Skinner; solo textos obtenidos por vía legal, sin Sci-Hub ni equivalentes; solo se cita lo que se leyó en el texto completo; los sitios que prohíben los rastreadores de IA, como los de Lupa y Aos Fatos (agencias brasileñas de verificación de datos), no se recolectan de forma masiva. La lista de candidatas vino en parte de nosotros (27 artículos de nuestro acervo y 12 libros nuestros) y en parte de la IA (97 artículos previstos a partir del índice y 34 añadidos después de 20 búsquedas dirigidas en OpenAlex). Aprobamos el relevamiento el 10 de octubre de 2026, y llegó a 168 fuentes, de las cuales 108 se leyeron en el texto completo y 60 quedaron pendientes (Apéndice D.2 y D.4).
- **Los criterios de análisis.** Lo que cuenta como contingencia, como consecuencia social y como control verbal. Clasificar un operante verbal (mando, tacto, intraverbal, tacto distorsionado) es siempre interpretación, y nunca dato. El mismo método vale para políticos de todos los campos; "falso" es siempre la clasificación de una agencia nombrada; "discurso de odio" es una etiqueta de una instancia externa; dato, interpretación e hipótesis van separados. Y lo que no pretendemos afirmar: no ofrecemos datos robustos sobre el fenómeno, sino lecturas que puedan orientar experimentos futuros en Análisis de la Conducta.
- **La estructura.** Los diez temas del pedido original (datos y estudios correlacionales, agencias de verificación, operantes verbales, equivalencia de estímulos, burbujas digitales, voto, economía de la atención, ciclo productor–audiencia, infodemiología e intervenciones), el capítulo sobre la ventana de Overton, que añadimos el 11 de octubre, y la aprobación del índice en cuatro partes, el 10 de octubre.

Lo que hizo la IA

Usamos Claude Sonnet 5.5, de Anthropic, en Claude Code, entorno en el que el modelo lee y escribe los archivos del proyecto, ejecuta scripts y consulta bases bibliográficas abiertas, entre el 10 y el 26 de octubre de 2026. El historial de versiones muestra también a Claude Opus 5.5 en tareas de organización del

repositorio y de diagramación, sin redacción de texto del libro. Las herramientas trabajaron siempre a partir del marco interpretativo y de las instrucciones que proporcionamos, y quien las condujo, de principio a fin, fue Márcio Moreira.

- **Relevamiento y obtención de las fuentes.** Propuso candidatas, comprobó autores, año, revista, volumen y páginas en Crossref, obtuvo los textos por vías legales (acceso abierto, repositorios, versiones puestas en línea por los autores y los PDF que adjuntamos) y extrajo el texto, con reconocimiento óptico de caracteres cuando la capa de texto era mala. Lo que no obtuvo quedó en una lista de pendientes y no se cita como verificado.
- **Estructura.** Propuso el índice detallado y el orden de escritura: los capítulos 4, 9, 6, 8 y 10 primero, porque tenían las fuentes completas, y los demás a continuación.
- **Redacción inicial.** A partir del marco, desarrolló las primeras versiones de todas las secciones teóricas, de la Presentación y de los Apéndices A y B.
- **Aplicación del marco a los estudios.** Produjo las primeras lecturas skinnerianas de los estudios seleccionados: descripción de los procedimientos, identificación de posibles contingencias, interpretaciones rivales y límites, y la interpretación de los resultados, siempre con dato, interpretación e hipótesis separados.
- **Verificación de las afirmaciones.** Registró cada afirmación atribuida a una fuente en una planilla, con las frases literales que la sostienen; un script buscó esas frases en el texto completo, y la IA leyó el contexto y registró el resultado y las salvedades. Un segundo script comprobó que toda cita del texto tiene referencia y que toda afirmación marcada existe y está verificada. La planilla tiene 1.207 entradas (Apéndice D.1). En el relevamiento inicial, la comprobación corrigió autores, años de volumen y páginas que la IA había registrado de memoria e hizo retirar una afirmación atribuida a un artículo que no la hace.
- **Figuras, referencias y archivos.** Generó las figuras por script y recortó, de los libros del Instituto Walden4, las imágenes de los estudios que describe el Capítulo 4; armó las referencias en APA y los archivos de entrega (.docx, HTML y PDF).
- **Reescrituras a pedido.** Reescribió en prosa las descripciones de estudios y rehízo el Capítulo 4 con las imágenes de nuestros estudios y de los cómics, según los pedidos descritos más abajo.

La IA no es autora. Decidió, por delegación nuestra, muchas cuestiones de redacción (el orden de los capítulos, la numeración, el título de capítulos, el tratamiento de las lagunas, lo que quedó fuera cuando una fuente no pudo obtenerse), y cada decisión está registrada con fecha y con el camino para revertirla. Tres términos que aparecen como definiciones propias del libro nacieron en la redacción con IA: "evitación del contracontrol" (Capítulo 6), "ventana de una audiencia" (Capítulo 10) y la distinción entre alcance, exposición y reacción (Capítulo 11). Cada uno se declara como propuesta en el texto, y su evaluación es nuestra. La palabra final sobre cada interpretación es de los autores.

La redacción se dio en sesiones de conversación con la IA sobre los archivos del proyecto, un capítulo a la vez, a partir del 10 de octubre de 2026. Comenzó por el pedido original, que el repositorio guarda sin alteraciones. La IA entregó por etapas (el artículo base, el índice, el relevamiento bibliográfico y solo entonces los capítulos), y aprobamos cada etapa antes de la siguiente. En otros momentos respondimos "puedes decidir por mí" y "Continúa la redacción del libro y finalízala. Puedes tomar todas las decisiones": la IA decidió y registró cada decisión, con fecha, en un archivo propio, y la delegación no valió como aprobación de capítulo. Pedimos también correcciones de rumbo después de leer el borrador: que el texto dejara de tratar a "el libro" como quien lee, dice y propone; que las aperturas de sección se sostuvieran sin el título; que las descripciones de investigaciones, truncadas y poco didácticas, se revisaran párrafo por párrafo; y que el Capítulo 4 se rehiciera con figuras en color y didácticas y con las imágenes de nuestros estudios, incluidas las de los cómics. Algunos de esos pedidos aparecen literalmente en la sección siguiente. El registro completo, con el pedido original, las decisiones fechadas, la planilla de verificación y el historial de versiones, está en el repositorio del proyecto y puede consultarse previa solicitud.

Ejemplos de prompts

Casi todos los pedidos fueron mensajes cortos, en lenguaje corriente, hechos en medio del trabajo; la excepción es el pedido original, largo y estructurado, que reproducimos íntegramente. Todos fueron escritos por Márcio Moreira. Los prompts se escribieron en portugués y se reproducen aquí traducidos. Los demás ejemplos son literales y provienen de distintos puntos del proceso; en ellos la ortografía está corregida, y las palabras añadidas o cambiadas están entre corchetes. Cada ejemplo va seguido de lo que la IA hizo con él, según el registro del proyecto.

El pedido original (10 de octubre de 2026), íntegro. Los títulos de las partes del pedido aparecen en negrita.

> **PAPEL**
>
> Eres un analista de la conducta (marco skinneriano, Análisis Experimental y Aplicado de la Conducta) con experiencia en conducta verbal, equivalencia de estímulos y análisis de prácticas culturales, que escribe para un público de grado avanzado y posgrado en Psicología.
>
> **OBJETIVO**
>
> Escribir el libro "Uma análise operante de discursos de ódio e desinformação", en portugués de Brasil, con lenguaje didáctico y rigurosamente técnico (terminología del Análisis de la Conducta, definiciones operacionales, sin mentalismo explicativo). El foco son los discursos y las declaraciones falsas de políticos brasileños.
>
> **FLUJO DE TRABAJO (ejecuta por etapas y detente a esperar mi aprobación al final de cada una)**
>
> 1. Lee primero el artículo base: https://pmc.ncbi.nlm.nih.gov/articles/PMC2733686/ y resume cómo fundamentará el análisis de la conducta verbal.
> 2. Propón el índice detallado (capítulos, secciones, figuras previstas, fuentes previstas).
> 3. Arma el relevamiento bibliográfico (ver "Reglas de fuentes") y la tabla de verificación, antes de escribir cualquier capítulo.
> 4. Escribe un capítulo a la vez, con figuras, y entrégalo como documento.
> 5. Al final, haz la revisión cruzada de todas las citas y referencias.
>
> **CONTENIDO (añade temas pertinentes que yo no haya enumerado)**
>
> a) Datos y estadísticas sobre desinformación ("fake news") y discurso de odio, incluidos estudios correlacionales entre propensión a la desinformación y posicionamiento político. Presenta las limitaciones (correlación ≠ causa, muestras, medidas, contexto cultural).
>
> b) Datos de agencias de verificación (Aos Fatos, Lupa, etc.) sobre declaraciones falsas o engañosas de Jair Bolsonaro durante el mandato y de Lula. Usa el mismo criterio para los dos: misma agencia, período comparable, misma metodología de clasificación, denominador explícito (n.º de declaraciones verificadas) y salvedad sobre el sesgo de selección de las verificaciones.
>
> c) Análisis de los discursos en términos de los operantes verbales (mando, tacto, ecoica, intraverbal, textual, etc., y tacto distorsionado/mando disfrazado), con base en el artículo base.
>
> d) Análisis en términos de equivalencia de estímulos: formación de clases, transferencia de función, discriminaciones condicionales, propiedades (reflexividad, simetría, transitividad, equivalencia), relación con la formación de prejuicio y actitudes. Referencia los libros de Márcio Moreira sobre actitudes y sobre cómics y actitudes (walden4.com.br/livros).

e) Burbujas digitales como comunidades verbales; creación de "realidades virtuales" por control verbal; evitación del contracontrol.

f) Conducta de votar como conducta verbal.

g) Economía de la atención; investigaciones sobre la tendencia de los usuarios a consumir contenido emocional/moralizado; algoritmos como programas de reforzamiento.

h) Ciclo de reforzamiento entre productor de contenido y audiencia ("me gusta", compartidos, audiencia como reforzadores generalizados), análisis de las contingencias de ambos.

i) Epidemiología de la información (infodemiología): difusión, tasas de propagación, intervenciones.

j) Implicaciones prácticas: intervenciones culturales y educativas basadas en el análisis de la conducta.

FIGURAS

Ilustra todo procedimiento o concepto que se beneficie de un diagrama: discriminaciones condicionales (MTS), clases de equivalencia y sus propiedades, triple contingencia, ciclo productor–audiencia, burbujas como comunidades verbales. Genera las figuras como archivos (Graphviz/matplotlib/SVG), con leyenda numerada, vectoriales, legibles en impresión en blanco y negro y originales (no reproducir figuras de terceros).

REGLAS DE FUENTES

- Cita solo artículos científicos, más sus libros sobre actitudes y cómics (excepción explícita). No cites otros libros.
- Obtén los artículos solo por vías legales: acceso abierto (PMC, SciELO, Unpaywall, OpenAlex, repositorios), acceso institucional que yo ponga a disposición, o PDF que yo adjunte. No uses Sci-Hub ni equivalentes. Lo que no se obtenga legalmente entra en una lista "PDF pendiente" y NO se cita como verificado.
- Para cada artículo citado, lee el texto completo y confirma que contiene la información atribuida. Regístralo en una tabla: referencia | DOI | fragmento/página que sustenta la afirmación | estado (verificado / pendiente).
- Valida DOI, autores, año, revista, volumen y páginas (p. ej., Crossref). Nunca inventes ni completes una referencia de memoria; si no puedes verificarla, márcala como pendiente.
- Haz dos pasadas de verificación (citas en el texto ↔ lista de referencias ↔ tabla de verificación). Usa el formato APA 7.
- Citas literales: cortas, entre comillas, con página, y acompañadas de una imagen (recorte) del fragmento del PDF original. Prefiere la paráfrasis fiel.

NOTICIAS

> Usa notas periodísticas solo de G1, CNN Brasil, BandNews y Canal Meio. Para cada una: fragmento corto citado con crédito, URL, fecha de publicación y fecha de acceso, más captura de pantalla del fragmento. Registra también la fecha de cada declaración analizada y la verificación correspondiente.
>
> **TONO Y RIGOR POLÍTICO**
>
> Describe conductas y contingencias, no "esencias" de personas o partidos. Analiza a políticos de diferentes campos con el mismo método. Separa siempre dato verificado, interpretación analítica e hipótesis. Evita afirmaciones difamatorias; atribuye cada clasificación de falsedad a la agencia que la hizo.
>
> **ENTREGA**
>
> Cada capítulo como documento (.docx), figuras en una carpeta aparte, tabla de verificación de referencias en una planilla, y lista de pendientes (PDF no obtenidos, datos no confirmados).

Lo que hizo la IA. De ese pedido vinieron el trabajo por etapas (artículo base, índice, relevamiento bibliográfico, capítulos y revisión cruzada), la planilla de verificación, la regla de citar solo lo que se leyó en el texto completo y la separación entre dato, interpretación e hipótesis en todos los capítulos. La instrucción de figuras legibles en blanco y negro se revisó después, a pedido nuestro: las figuras del Capítulo 4 pasaron a ser en color el 16 de octubre, y las demás el 26 de octubre de 2026.

Aprobar y seguir (10 de octubre de 2026). Entre las etapas, respondimos con frases como estas:

> Adelante.
>
> Aprobado. Continúa.

Lo que hizo la IA. Trató cada respuesta como aprobación solo de la etapa que cubría la pregunta; los puntos que no se respondieron siguieron abiertos en el registro de decisiones.

Delegar decisiones (10 y 14 de octubre de 2026).

> Puedes decidir por mí.
>
> Continúa la redacción del libro y finalízala. Puedes tomar todas las decisiones.

Lo que hizo la IA. Decidió lo que faltaba (orden de escritura, numeración, tratamiento de las fuentes que no se obtuvieron, títulos) y registró cada decisión, con fecha y con el camino para revertirla. La delegación no valió como aprobación de capítulo.

Ampliar el alcance (11 de octubre de 2026).

> Añade un capítulo con un análisis conductual de la ventana de Overton, relacionándolo con discursos de políticos y redes sociales.

Lo que hizo la IA. Escribió el capítulo, hoy el Capítulo 10, y propuso la definición de "ventana de una audiencia" que usa, declarada como propuesta en el texto.

Responder a una pregunta de la IA, con una respuesta ambigua (11 de octubre de 2026). La IA había preguntado si la excepción de fuentes abierta para Skinner valía también para la obra de Ferster y Skinner sobre programas de reforzamiento. La respuesta fue:

> Fester, ok. Adelante.

Lo que hizo la IA. Leyó "Fester" como Ferster y Skinner, registró esa lectura y avisó que, si el libro pretendido era otro, bastaba con decirlo.

Corregir la redacción después de leer el borrador (16 de octubre de 2026).

> "El libro" entr[a] como una entidad en el propio libro (e.g., "El primer estudio que el libro leyó es un juego"). Hay párrafos de inicio de sección que empiezan de forma abrupta y confusa o considerando que el título de la sección ya [forma] parte del párrafo. Identifica y ajusta esas cuestiones.

Lo que hizo la IA. Reescribió unas 185 frases en voz impersonal, en todos los capítulos, y rehízo las aperturas de sección para que se entiendan sin el título.

> Hay descripciones de investigaciones que están truncadas, poco didácticas. Revisa párrafo a párrafo esa cuestión de la redacción.

Lo que hizo la IA. Reescribió en prosa las descripciones de estudios de los doce capítulos y de la Presentación, una por una, con la pregunta, la muestra, el procedimiento, la medida, el resultado y la salvedad de cada estudio.

Pedir detalle e imágenes (16 de octubre de 2026).

> En el capítulo de actitudes y equivalencia, usa en detalle y usando las imágenes de estos dos trabajos: [direcciones de los dos trabajos]. Usa figuras en color, muy didácticas. Rehaz el capítulo con este tono.
>
> Inserta, donde sea posible y adecuado, referencias a: [direcciones de once trabajos]. Usa imágenes de esos libros, deja todo bien ilustrado, incluidas imágenes de los cómics.

Lo que hizo la IA. Rehízo el Capítulo 4 con figuras en color, recortó imágenes de los estudios y de los cómics del Instituto Walden4 e incluyó las obras en las referencias, con la dirección de cada PDF. Una de las direcciones aparecía dos veces en la lista, y la IA la leyó como el volumen siguiente, registrando la suposición para que la confirmáramos.

Lo que hicimos después

Entre el 19 y el 26 de octubre de 2026, los autores trabajaron sobre el texto producido en las etapas anteriores: revisaron, modificaron, ajustaron y comprobaron los capítulos, la Presentación y los apéndices, en un manuscrito común.

1. **Revisión conceptual.** Evaluamos si cada argumento era coherente con el marco skinneriano y con el análisis contextual, y descartamos o rehicimos los que no lo eran.
2. **Modificación y ajuste.** Corregimos errores conceptuales, inferencias sin apoyo en los datos y extrapolaciones, y reescribimos los fragmentos que pedían otra formulación.
3. **Comprobación.** La comprobación de cada afirmación contra el texto completo de la fuente la hizo primero la IA, con el procedimiento del Apéndice D; nosotros comprobamos los textos en esa misma semana de revisión. Las 60 fuentes que no pudimos obtener no se citan como verificadas, y donde el asunto exigiría la fuente el capítulo dice que no se leyó.
4. **Edición final.** Unificamos terminología, estilo y la articulación entre capítulos.

El recuadro siguiente resume las etapas por capítulo.

RECUADRO

Etapas del trabajo, por capítulo

Capítulo	Primera versión por la IA	Obras en la lista del capítulo	Entradas verificadas	Revisión y comprobación por los autores
Presentación	15/10/2026	2	0	19 a 26/10/2026
1. Definiciones y datos	14 a 15/10/2026	21	48	19 a 26/10/2026
2. Agencias de verificación	10/10/2026	10	39	19 a 26/10/2026
3. Declaraciones falsas	14 a 15/10/2026	13	92	19 a 26/10/2026
4. Actitudes y equivalencia	10/10/2026, rehecho el 16/10/2026	39	300	19 a 26/10/2026
5. Discurso de odio	14 a 15/10/2026	13	45	19 a 26/10/2026
6. Burbujas digitales	11/10/2026	13	87	19 a 26/10/2026
7. El voto	14 a 15/10/2026	13	54	19 a 26/10/2026
8. Economía de la atención	11/10/2026	12	80	19 a 26/10/2026

Capítulo	Primera versión por la IA	Obras en la lista del capítulo	Entradas verificadas	Revisión y comprobación por los autores
9. Ciclo productor–audiencia	10/10/2026	13	123	19 a 26/10/2026
10. Ventana de Overton	11/10/2026	11	80	19 a 26/10/2026
11. Infodemiología	14 a 15/10/2026	12	97	19 a 26/10/2026
12. Intervenciones	14 a 15/10/2026	27	162	19 a 26/10/2026

Nota. Las fechas de la primera versión son las del primer registro de cada archivo en el historial del proyecto; el Capítulo 10 se comenzó el 11/10/2026 con otro número y se renumeró el 14/10/2026. “Entradas verificadas” son las afirmaciones y los datos de fuentes registrados en la planilla de verificación para el capítulo, cada uno con el fragmento leído en contexto, antes de la revisión de los autores y de la sección 6.7. El 16/10/2026 pedimos para todos los capítulos dos revisiones de redacción (el texto sin “el libro” como agente y las descripciones de estudios en prosa), y el Capítulo 4 se rehízo. Los Apéndices C y D se generan por script a partir de la planilla.

Límites del proceso

Dos límites merecen mencionarse. El primero es que la verificación confirma que el pasaje atribuido a una fuente está en el texto de esa fuente, en la página indicada, y no que la lectura operante que se hace de él sea correcta; esa es la parte que los capítulos llaman interpretación o hipótesis, y la que más depende de nuestro juicio. El segundo es que los modelos de lenguaje se equivocan de memoria, y la propia IA registró, al inicio del trabajo, datos que Crossref y el texto de las fuentes desmintieron; por eso nada entró en el libro sin la comprobación descrita en el Apéndice D, y donde la fuente no pudo leerse el texto dice que no se leyó.

Una lectura conductual del propio proceso

Para los analistas de la conducta, la pregunta sobre la autoría tiene una respuesta conocida, y no cambia a causa de la IA. En un ensayo de 1972, “On ‘having’ a poem”, Skinner compara escribir un poema con tener un hijo: el poeta no crea a partir de la nada, es el locus donde se encuentran variables genéticas y ambientales. Lo que llamamos creatividad es variación y selección, y buena parte del trabajo del escritor está en editar, es decir, en seleccionar lo que permanece.

No invocamos a Skinner para disminuir el papel de la IA ni para inflar el nuestro. La analogía sirve para describir lo que de hecho cambió: la IA amplió mucho la fuente de variación, y el trabajo autoral se concentró aún más en la disposición de las condiciones (el problema, el marco, las instrucciones) y en la

selección (lo que se aceptó, corrigió, reescribió o descartó). Esta es una descripción del proceso, no una justificación de él; si el lector discrepa de ella, la descripción de las secciones anteriores sigue valiendo por sí misma.

Responsabilidad

Los autores revisaron el contenido de este libro y asumen la responsabilidad por él, incluidas las partes cuya primera redacción se hizo con IA. Los errores que queden son nuestros.

La edición en español

Esta edición se tradujo de la edición brasileña en portugués (ISBN 978-85-65721-74-5) con Claude (Anthropic), en Claude Code, bajo la dirección de Márcio Borges Moreira, que firma como traductor. La traducción partió del texto portugués. De la edición en inglés se tomaron las correcciones que los autores aprobaron para ella: el nombre y el año de una comisión parlamentaria, años de referencias, remisiones internas, un recuento, algunas frases incompletas o ambiguas del original, la explicación del signo de un tamaño del efecto y dos figuras. Las citas literales de obras escritas en inglés son traducciones nuestras, hechas a partir de las palabras originales cuando la edición en inglés las había recuperado o cuando constaban en los registros de verificación del libro, y del portugués en los demás casos; no se consultaron traducciones publicadas. Las citas de fuentes en portugués también son traducciones nuestras. Las figuras se redibujaron a partir de los mismos datos, con rótulos en español; las imágenes tomadas de los estudios y de los cómics del Capítulo 4 y las portadas de las publicaciones del Capítulo 6 conservan el texto en portugués, y las notas de las figuras del Capítulo 4 lo señalan. Otra instancia de inteligencia artificial comparó después la traducción, frase por frase, con el original (revisión adversarial), y se corrigieron los errores encontrados. Los ejemplos, las instituciones y las agencias de verificación brasileñas se mantuvieron, con una breve glosa la primera vez que aparecen.

instituto WALDEN4
walden4.com.br



PARTE I

El fenómeno y los datos

1 Desinformación y discurso de odio: definiciones, medidas y datos

2 Lo que dicen las agencias de verificación (y lo que no pueden decir)

CAPÍTULO

1

Desinformación y discurso de odio: definiciones, medidas y datos

Las cifras de Allcott y Gentzkow (2017) sobre la elección presidencial estadounidense de 2016 muestran, desde el principio, el problema que organiza este capítulo. En la base de noticias falsas que reunieron, 115 favorables a Trump se compartieron 30 millones de veces en Facebook, y 41 favorables a Clinton, 7,6 millones. Los mismos autores estiman que el adulto estadounidense promedio vio y recordó 1,14 noticias falsas y dicen, de forma expresa, que no evalúan si estas decidieron la elección. Un lector apresurado junta esas cifras en una frase, y estas responden a preguntas diferentes: la primera mide lo que circuló; la segunda, lo que alguien vio y recordó; ninguna mide lo que cambió en lo que alguien hizo.

Este capítulo sistematiza el conocimiento acumulado sobre la cantidad y la forma de circulación de la desinformación y del discurso de odio, así como el perfil de quienes los producen y comparten, explicitando las métricas empleadas y sus límites metodológicos. Es el capítulo menos analítico del libro: no propone una lectura operante, sino que muestra los datos con que trabajan los capítulos siguientes y lo que cada uno de ellos exige que se diga en voz baja. Vale en él la regla de todo el libro: antes de leer la cifra, preguntar qué controla el instrumento.

1.1 Definiciones, y lo que cada una exige medir

Cada definición de "información falsa" trae incorporada una exigencia de medida. Lazer et al. (2018) definen la noticia falsa como información fabricada que imita el contenido periodístico en la forma, pero no en el proceso organizacional ni en la intención; es una subcategoría de la desinformación en senti-

do amplio (misinformation), es decir, información incorrecta o engañosa sobre el estado del mundo, y sus vehículos carecen de las normas y los procesos editoriales de los medios. Si el elemento definitorio es la intención y el proceso del editor, escriben ellos, la prevalencia debe medirse por editor, y no por noticia individual. Allcott y Gentzkow (2017) definen la noticia falsa como un artículo de noticias intencional y verificablemente falso, que puede engañar al lector, y excluyen de la definición los errores no intencionales de reportaje. Las dos definiciones exigen, en principio, conocer la intención; ninguna de las medidas que las usan la observa.

Los estudios sobre compartir resuelven el problema cambiando el ítem por la fuente. Guess et al. (2019) toman como noticia falsa el contenido falso o engañoso vestido intencionalmente como noticia, en general para generar ingresos publicitarios, y cruzan los enlaces que cada encuestado compartió con listas de editores de noticias falsas hechas por periodistas y académicos. Lasser et al. (2022) usan la nota que NewsGuard, un servicio de evaluación de la confiabilidad de sitios de noticias, da al dominio (de 0 a 100; por debajo de 60, "no confiable") como estimador de la confiabilidad del ítem compartido. Es una convención explícita, la de inferir el ítem por la fuente, y no un defecto escondido; pero cambia lo que dice la cifra: la proporción de enlaces a dominios no confiables no es la proporción de afirmaciones falsas.

Los estudios experimentales cambian de nuevo el objeto. Rathje et al. (2023) escriben que misinformation puede referirse a noticias fabricadas, rumores falsos, teorías de la conspiración o desinformación, y lo que miden es el juicio que los participantes hacen de titulares verdaderos y falsos. Y Ricard y Medeiros (2020), en una nota sobre el caso brasileño, distinguen la información falsa que quien la disemina cree verdadera de aquella que quien la disemina sabe que es falsa —dos puntos de un "espectro del 'desorden informativo'", sin presentar una medición empírica propia. La distinción es de intención, y la Tabla 1.1 muestra que casi ninguna de las medidas la observa.

Los términos se usan como las fuentes los traen. "Información falsa" y "desinformación" aparecen, sin pretensión de definición propia, para lo que la fuente llama misinformation, y la intención solo entra cuando la fuente la mide. En cuanto a "discurso de odio", la regla es la del resto del libro: es un rótulo externo, aplicado solo al enunciado clasificado por una instancia externa, con la instancia nombrada; los estudios de la sección 1.7 usan la expresión como categoría de la propia medida, y aquí se toma como ellos la traen.

TABLA 1.1

Seis definiciones de información falsa y lo que cada una exige medir

Fuente	Término y definición	Lo que la definición exige medir	Lo que el estudio midió
Lazer et al. (2018)	Noticia falsa: información fabricada que imita la forma del periodismo, no el proceso ni la intención	Proceso e intención del editor	Sin medida propia: ensayo que propone medir por editor, y no por noticia
Allcott y Gentzkow (2017)	Noticia falsa: artículo intencional y verificablemente falso, que puede engañar	Intención y veracidad, noticia por noticia	Veces que se compartieron 156 noticias clasificadas como falsas por verificadores en los tres meses previos a la elección
Guess et al. (2019)	Contenido falso o engañoso vestido como noticia, en general por ingresos publicitarios	Origen del contenido	Enlaces compartidos en Facebook, cruzados con listas de editores de noticias falsas
Lasser et al. (2022)	Información no confiable, definida por el dominio	Confiabilidad de la fuente, no del ítem	Nota de NewsGuard por debajo de 60 para el dominio del enlace
Rathje et al. (2023)	Misinformation: noticias fabricadas, rumores, teorías de la conspiración o desinformación	Veracidad del titular	Juicio de titulares verdaderos y falsos por los participantes, con y sin incentivo
Ricard y Medeiros (2020)	Misinformation (quien disemina cree) y disinformation (quien disemina sabe), citando otra obra	Lo que sabe quien disemina	Sin medida propia: nota comentada, sin datos empíricos

Nota. Descripciones propias, a partir de los pasajes citados; la columna “Lo que el estudio midió” trae lo que cada estudio operacionalizó, y no lo que declaró querer medir.

1.2 Prevalencia: cuánto circula y quién comparte

La prevalencia puede investigarse a partir de tres preguntas centrales, y cada una exige un instrumento de medida propio: cuánto circula, quién comparte y de dónde proviene el contenido. La primera pregunta se refiere a cuánto circula. Allcott y Gentzkow (2017) armaron, para la elección presidencial estadounidense de 2016, una base con 156 noticias relacionadas con la elección que sitios de verificación de referencia habían clasificado como falsas en los tres meses anteriores a la votación. Las 115 favorables a Trump se compartieron 30 millones de veces en Facebook, y las 41 favorables a Clinton, 7,6 millo-

nes de veces. Compartir, sin embargo, no es lo mismo que haber visto, y los autores intentaron estimar lo que las personas vieron: en una encuesta en línea con 1.200 personas, hecha después de la elección, preguntaron si recordaban haber visto cada uno de 14 titulares falsos que habían circulado. Cerca del 15% dijo que sí, pero cerca del 14% también dijo haber visto titulares placebo, falsos, inventados por los autores y que nunca circularon. Como los titulares placebo no podían haber sido vistos, los autores toman la diferencia entre los dos porcentajes como medida de recuerdo verdadero y, proyectándola al conjunto de noticias de la base, estiman que el adulto promedio vio y recordó 1,14 noticias falsas. En la misma encuesta, solo el 14% de los adultos estadounidenses veía las redes sociales como la fuente más importante de noticias electorales.

La segunda pregunta es quién comparte, y el estudio de Guess et al. (2019) es el que más se acerca a una respuesta con registro de conducta, y no solo con autoinforme. Los autores aplicaron una encuesta en línea a una muestra representativa de 3.500 estadounidenses y pidieron a cada encuestado que autorizara el acceso, mediante una aplicación, a datos de su propio Facebook, entre ellos las publicaciones con enlaces externos; el análisis abarca a quienes dieron su consentimiento y tuvieron sus datos vinculados a la encuesta. Los enlaces compartidos en 2016 se compararon entonces con listas de editores de noticias falsas hechas por periodistas y académicos, lo que evita los sesgos conocidos del autoinforme de actividad en línea. El resultado principal es de rareza: cerca del 8,5% de los encuestados con datos vinculados compartió al menos una nota de un dominio de noticias falsas, y más del 90% no compartió ninguna. No es que las personas compartan poco en general: en el período, el 26,1% de los encuestados compartió de 10 a 100 enlaces de cualquier tipo, y el 61,3%, de 100 a 1.000. Entre quienes compartieron noticias falsas había más republicanos: 38 contra 17 demócratas en números absolutos, y el 18,1% de los republicanos contra el 3,5% de los demócratas en la muestra. El hallazgo más robusto, dicen los autores, es la edad: quienes tenían 65 años o más compartieron en promedio 0,75 notas de dominios de noticias falsas, más del doble que el segundo grupo de edad más avanzada (0,26), y la diferencia persiste cuando se mantienen constantes la ideología y la educación.

Dos salvedades de los propios autores acompañan estas cifras. La primera es que el patrón por ideología es coherente con dos hechos: la mayoría de las noticias falsas tenía una inclinación pro-Trump, y las personas tienden a compartir lo que aprueban. El dato, por lo tanto, puede no indicar que los conservadores tengan mayor tendencia a compartir noticias falsas cuando se exponen a ellas. La segunda es que los autores no tienen una medida de si los encuestados sabían que el contenido era falso (Guess et al., 2019). Lazer et al.

(2018) añaden un límite de método. Citan un estudio que evaluó la diseminación de un conjunto de noticias falsas destacadas y estimó que el estadounidense promedio se encontró con entre una y tres de ellas en el mes anterior a la elección; la estimación es probablemente conservadora, porque el estudio siguió solo 156 noticias. Cabe subrayar que cuantificar las cuentas que recibieron o replicaron una publicación no revela cuántas personas la leyeron, ni si ese estímulo ejerció algún efecto sobre sus acciones.

En síntesis, como formulan Lazer et al. (2018): exposición no equivale a impacto, y la evaluación, a mediano y largo plazo, del efecto de la exposición a noticias falsas sobre la conducta política es esencialmente inexistente en la literatura. Mosleh y Rand (2022) recuerdan una distinción parecida del lado de quien comparte: exposición y compartir son constructos fundamentalmente diferentes, la mayor parte de las personas comparte solo una pequeña fracción de lo que ve, y mirar lo que alguien comparte da un retrato muy limitado de su ambiente de información.

1.3 Las élites políticas como fuente

La tercera pregunta sobre la prevalencia —de dónde proviene la información falsa— desplaza el foco del público en general a las élites políticas. Mosleh y Rand (2022) argumentan que, después de la atención a las noticias de editores oscuros, una forma diferente de desinformación pasó a llamar la atención: campañas coordinadas por figuras y organizaciones públicas, a las que llaman élites. Para estudiarla, construyeron dos puntuaciones. La primera es una puntuación de falsedad para cada una de 816 élites, calculada a partir de las verificaciones profesionales de PolitiFact sobre la veracidad de sus declaraciones. La segunda es una puntuación de exposición para cada usuario de Twitter, calculada a partir de las puntuaciones de falsedad de las élites que sigue. Después compararon la exposición con lo que el propio usuario comparte. La calidad de las noticias compartidas se midió por la confiabilidad del dominio, evaluada de dos formas: por el promedio de ocho verificadores profesionales y por una muestra de 970 estadounidenses elegidos por cuotas, con las notas de simpatizantes de los dos partidos equilibradas. La ideología del usuario se estimó por la ideología de las cuentas que sigue. Cuanto mayor la exposición a la desinformación de élites, menor la calidad de las noticias que el propio usuario comparte y más a la derecha se sitúa su inclinación política estimada. Es una correlación entre medidas, y los propios autores dicen que el enfoque se basa en la red inmediata de cuentas seguidas y no capta la exposición más allá de ellas, como la que llega por retuits de conexiones o por recomendaciones algorítmicas.

Lasser et al. (2022) analizan directamente las publicaciones emitidas por parlamentarios. Los autores reunieron los tuits, sin retuits, de miembros del Congreso estadounidense, del parlamento alemán y del parlamento británico entre 2016 y marzo de 2022, extrajeron los enlaces y le dieron a cada uno la nota de NewsGuard para el dominio, de 0 (muy malo) a 100 (periodismo de calidad excepcional); por debajo de 60, el dominio se considera "no confiable". Los republicanos compartieron 9,1 veces más enlaces a dominios no confiables que los demócratas: el 3,86% de los enlaces, contra el 0,43%. En Alemania, la CDU/CSU (democristianos) compartió el 0,18% y el SPD (socialdemócratas), el 0,03%; en el Reino Unido, los conservadores el 0,25% y los laboristas el 0,05%, y los autores señalan que esas cifras se basan en recuentos muy bajos. La proporción de enlaces no confiables de los republicanos pasa del 2,4% en 2016–2018 al 5,5% en 2020–2022, mientras que la de los demócratas sigue en el 0,4%; el resultado se repite con una segunda base de confiabilidad de dominios, compilada de forma independiente de NewsGuard. La divergencia entre los partidos parece peculiar de los Estados Unidos: no aparece en Alemania ni en el Reino Unido.

Lo que los autores proponen como explicación es una posibilidad debatida: la motivación partidaria, ya que el contenido despectivo sobre el grupo político adversario aumenta las intenciones de compartir entre los partidarios. Los autores sugieren también que el patrón entre los parlamentarios puede contribuir a la asimetría que se observa en el público, de dos maneras: al proporcionar desinformación a los partidarios y al legitimar que se comparta información no confiable. NewsGuard solo tiene nota para una parte de los enlaces, y el estudio registra la cobertura: el 46,5% de los enlaces del Congreso estadounidense, el 58,8% del parlamento alemán y el 39,2% del británico, después de excluir las redes sociales y los buscadores. Lo que queda sin nota no se mide.

1.4 Posicionamiento político y propensión a la información falsa: lo que muestran los estudios correlacionales

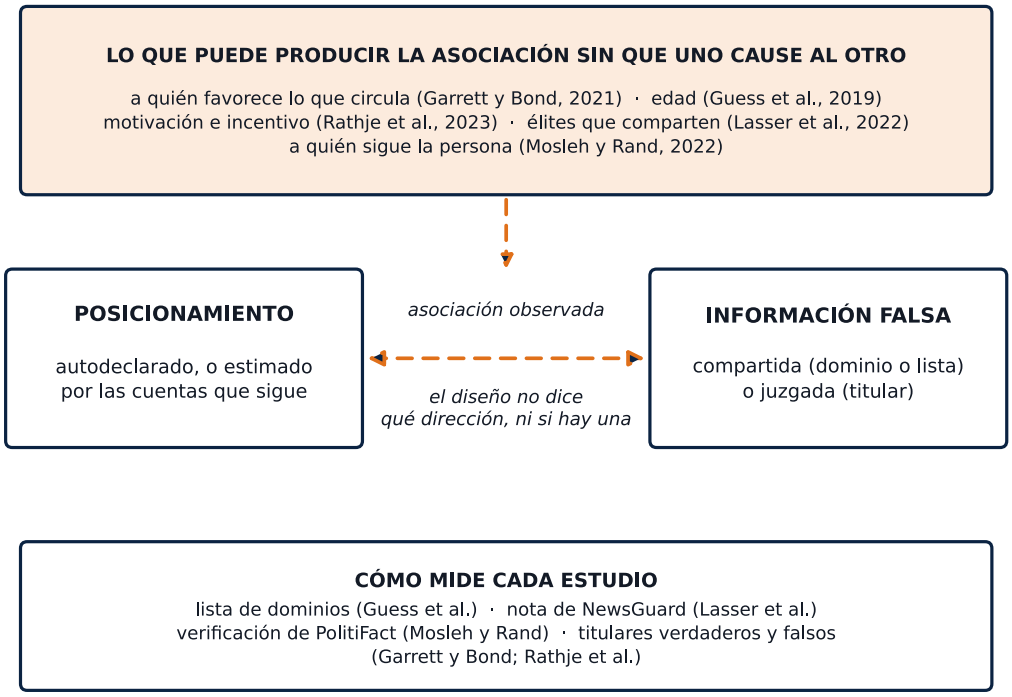
La pregunta recurrente en la literatura es si hay asimetría partidaria: si quien se posiciona en determinado campo comparte o cree más en información falsa. Los estudios verificados responden con asociaciones, y la Tabla 1.2 las reúne con lo que cada diseño no permite concluir. La Figura 1.1 ilustra los límites de esa inferencia.

Garrett y Bond (2021) adoptaron un cuidado metodológico poco común en el área. La evidencia de que los conservadores tienen más creencias falsas se apoyaba en un conjunto relativamente estrecho de temas, en muchos casos elegidos a propósito para reflejar falsedades promovidas por élites políticas

conservadoras. Los autores, en cambio, dejaron la elección de las afirmaciones a un criterio externo: cada dos semanas, un servicio identificó las 20 historias políticas más virales, 10 verdaderas y 10 falsas, y un panel de 12 oleadas, con una gran muestra representativa de estadounidenses, evaluó en la semana siguiente afirmaciones cortas sobre esas historias, a lo largo de seis meses. Con las respuestas, calcularon dos índices, mediante la teoría de detección de señales: la sensibilidad, que es la capacidad de distinguir afirmaciones verdaderas de falsas, y el sesgo de respuesta, que es la propensión a rotular todas las afirmaciones como verdaderas o como falsas. En datos de encuestas en los Estados Unidos, los conservadores distinguieron verdades de falsedades peor que los liberales, y los autores atribuyen parte de la diferencia al ambiente informativo: las afirmaciones verdaderas de alto perfil tienden a promover temas y candidatos favorecidos por los liberales, y las narrativas falsas tendían a favorecer causas conservadoras, de modo que, aun con igual sesgo, los liberales tendrían un mejor desempeño en ese ambiente.

FIGURA 1.1

Lo que muestra una asociación entre posicionamiento político e información falsa, y lo que puede producirla sin que uno cause al otro



Esquema original, sin datos. Discontinua: relación que los diseños correlacionales no separan.

Nota. Esquema original del autor, a partir de Garrett y Bond (2021), Guess et al. (2019), Rathje et al. (2023), Lasser et al. (2022) y Mosleh y Rand (2022). No representa datos. Discontinua: relación que los diseños correlacionales no separan.

Rathje et al. (2023) añaden experimentos en los que se pudo variar la consecuencia de acertar. En cuatro experimentos, con 3.364 participantes de los Estados Unidos, las personas leían titulares políticos verdaderos y falsos y respondían si la afirmación del titular era precisa; a una mitad, sorteada, se le informó que recibiría una bonificación de hasta un dólar según las respuestas correctas, y la otra mitad respondió a las mismas preguntas sin incentivo. Los autores llaman sesgo partidario a la diferencia entre el número de titulares favorables al propio partido y el de titulares contrarios al propio partido que el participante juzgó verdaderos. Los incentivos financieros por exactitud redujeron el sesgo partidario en el juicio de titulares en cerca de un 30%. Los conservadores fueron menos exactos que los liberales al distinguir titulares verdaderos de falsos, pero los incentivos redujeron la diferencia en un 52%; e incentivar a los participantes a identificar noticias que agradarían a sus aliados políticos disminuyó la exactitud. Parte de lo que aparece como diferencia de creencia, por lo tanto, parece depender de lo que se refuerza: acertar o agradar.

TABLA 1.2

Cinco estudios sobre posicionamiento político e información falsa: lo que midieron y lo que no permiten concluir

Estudio	Muestra y país	Medida de información falsa	Medida de posicionamiento	Resultado	Lo que no permite
Guess et al. (2019)	Encuestados de una encuesta representativa vinculados a datos de Facebook; EE. UU., 2016	Enlaces de dominios de noticias falsas, por lista	Autodeclarado (partido, ideología)	Raro (8,5%); más entre republicanos y entre los de más edad; la edad es el hallazgo robusto	Atribuir la diferencia a una mayor propensión: puede reflejar la oferta pro-Trump
Garrett y Bond (2021)	Encuesta en EE. UU., afirmaciones a lo largo de 6 meses	Juicio de afirmaciones verdaderas y falsas	Autodeclarado	Los conservadores discriminaron peor; una parte se atribuye al ambiente informativo	Concluir que hay un sesgo mayor: el resultado depende del conjunto de afirmaciones

Estudio	Muestra y país	Medida de información falsa	Medida de posicionamiento	Resultado	Lo que no permite
Mosleh y Rand (2022)	816 élites y usuarios de Twitter, EE. UU.	Puntuación de falsedad según verificaciones de PolitiFact	Estimado por las cuentas que sigue el usuario	La exposición a élites con alta puntuación de falsedad se correlaciona con peor calidad de noticias y con ideología conservadora	Causalidad; un solo verificador; exposición más allá de las cuentas que se siguen
Lasser et al. (2022)	Parlamentarios de EE. UU., de Alemania y del Reino Unido, 2016–2022	Nota de NewsGuard del dominio del enlace	Partido	Los republicanos comparten más, y más con el tiempo; Alemania y Reino Unido, recuentos bajos	Inferir la veracidad de la declaración a partir del dominio de la fuente; cobertura restringida del 39% al 59% de los enlaces
Rathje et al. (2023)	Cuatro experimentos, n = 3.364, EE. UU.	Juicio de titulares verdaderos y falsos	Autodeclarado	El incentivo redujo el sesgo partidario en cerca de un 30% y la diferencia en un 52%	Generalizar a partir de titulares y participantes en línea

Nota. Datos compilados a partir de los artículos originales citados. Las restricciones de la última columna explicitan lo que los diseños metodológicos no logran aislar.

1.5 Limitaciones: lo que los diseños no permiten

Los estudios reunidos en las secciones anteriores tienen tres límites en común, y es a partir de ellos que deben leerse las cifras del capítulo. De los cinco estudios de la Tabla 1.2, cuatro son sobre los Estados Unidos y el quinto compara con Alemania y el Reino Unido; los datos de tres de ellos vienen de una plataforma, Facebook o Twitter. El primer límite es, por lo tanto, de muestra y de contexto: los hallazgos son del sistema bipartidista estadounidense, en elecciones específicas, y Lasser et al. (2022) dicen que la divergencia que observan parece peculiar de los Estados Unidos. El segundo es de causa: ninguno de esos diseños separa lo que la persona ve, lo que se le ofrece y lo que ella elige, y la Figura 1.1 reúne las variables que pueden producir la asociación

sin que posicionamiento e información falsa se causen. El tercero es de medida, y merece un recuadro, porque es lo que se toma prestado del análisis de la conducta para leer todos los demás.

CONCEPTO EN USO

Conducta verbal, audiencia y autoinforme

Conducta verbal es la conducta reforzada por la mediación de otras personas (Skinner, 1957); la audiencia es el estímulo discriminativo en cuya presencia esa conducta es característicamente reforzada. El autoinforme, es decir, responder a una escala, a una pregunta de encuesta o a una entrevista, es conducta verbal, y está bajo control de quien pregunta y de cómo se hace la pregunta. Para Guerin (1994), las escalas y las preguntas de encuesta dependen de quién pregunta y de la cultura verbal en que se hace la encuesta, porque las actitudes dependen siempre de consecuencias sociales. El ejemplo clásico es el de Hildum y Brown (1956): en un sondeo telefónico con 40 estudiantes varones de Harvard, el entrevistador decía "Good" después de las respuestas a favor, o en contra, del programa de Educación General, y el "Good" desplazó las respuestas en la dirección reforzada; solo 8 de 20 sujetos notaron lo que hacía el entrevistador, y todos rechazaron que sus respuestas hubieran sido influidas por él.

La consecuencia para este capítulo es directa. La Tabla 1.2 trae tres filas en las que el dato es una respuesta a una encuesta o a una tarea de juicio (Garrett y Bond, Rathje et al. y, en parte, Guess et al., donde la ideología es autodeclarada), y una respuesta a una encuesta es conducta bajo control de la situación en que fue dada. El resultado de Rathje et al. ilustra precisamente esa dinámica: cuando la consecuencia de acertar cambia, el juicio verbal se altera. Un segundo problema de medida es el de la selección de los estímulos. Cameoka y Moreira (2021) relatan tres estudios con universitarios que evaluaron fotos de hombres blancos y de hombres negros. En el Estudio 1, 75 participantes evaluaron, en una escala de 1 (malo) a 10 (bueno), 10 fotos de hombres blancos y 10 de hombres negros con expresiones neutras; el Estudio 2 tuvo 40 participantes, con rostros ligeramente sonrientes, y el Estudio 3, 30. La diferencia entre las medias de los dos grupos de fotos fue significativa en los Estudios 1 y 2 y no en el 3, en el que se usaron las fotos de blancos con las evaluaciones más altas y las de negros con las más bajas en los dos primeros estudios. Los autores toman el Estudio 3 como demostración del potencial de distorsión de una selección de estímulos poco rigurosa y sugieren que la selección de los estímulos puede sesgar la medición. Lo que se mide depende de lo que se eligió para ser medido.

Queda además el límite del instrumento de verificación, que el Capítulo 2 trata por entero: la proporción de falsas entre las declaraciones verificadas no es la proporción entre todas las declaraciones, y una agencia no es un sensor

neutro. Mosleh y Rand (2022) usan un único verificador, PolitiFact, y élites de los Estados Unidos. El lector debe guardar, para los demás capítulos, lo que esta sección establece: toda tasa de falsedad es función de quién fue medido, con qué instrumento, en qué situación, y ninguna de ellas es, por sí misma, la tasa de un campo político.

1.6 Lo que se sabe sobre Brasil

Lo que las fuentes verificadas dicen de Brasil es menos de lo que la sección 1.2 dice de los Estados Unidos, y de naturaleza diferente. Ricard y Medeiros (2020), en un comentario, y no en un estudio empírico, escriben que una retórica de negación de los riesgos de la covid-19 fue respaldada en los más altos niveles del gobierno brasileño y apoyada por desinformación; y registran que, según la comisión parlamentaria mixta de investigación sobre las noticias falsas (la Comissão Parlamentar Mista de Inquérito, o CPMI das Fake News, instalada por el Congreso Nacional de Brasil en 2019), parte de la desinformación sobre la pandemia atacaba a responsables de decisiones y figuras públicas, como gobernadores, los medios y especialistas en salud, para deslegitimar a quienes apoyaban el aislamiento, y que la investigación aún no había divulgado evidencia de coordinación directa por el llamado Gabinete do Ódio ("Gabinete del Odio", grupo de asesores presidenciales acusado de campañas de difamación en línea) o por miembros del gobierno.

Carvalho et al. (2021) hicieron un análisis de contenido, con el programa IRaMuTeQ, de publicaciones en los perfiles de Trump y de Bolsonaro en Facebook y de pronunciamientos oficiales en los sitios de los dos gobiernos, desde el primer día de notificación de la covid-19 en cada país. Los autores señalan similitudes temáticas y posiciones contrarias a las medidas defendidas por la OMS. Escriben que no es posible afirmar categóricamente que hubo una alineación intencional entre los dos, pero que el análisis muestra una convergencia en las estrategias de desinformación, y que la minimización de los riesgos de contagio, sobre todo en las publicaciones de Facebook, es una de las marcas de la similitud. Es un análisis de dos emisores, y no un dato de prevalencia.

Recuero et al. (2025) analizaron 811 verificaciones de tres agencias brasileñas sobre las elecciones de 2018 y 2022, mediante análisis de contenido y métodos mixtos, comparando temáticas, tipos de desinformación y las fuentes que legitimaban el contenido. Señalan una transición de contenidos totalmente falsos a engañosos, y de contenidos sin fuentes a contenidos que usan fuentes reales; leen el cambio de plataforma de los contenidos verificados, de Facebook a WhatsApp y Twitter entre 2018 y 2022, como una adaptación de la desinforma-

ción a espacios más cerrados, en los que es más difícil que circulen contradiscursos. Registran, igualmente, que la literatura revisada señala un mayor volumen de desinformación generada por partidarios de Bolsonaro en ambas elecciones, aunque ambas campañas hayan figurado a la vez como polos emisores y como blancos de desinformación. Lupa (Agência Lupa, 2026), agencia brasileña de verificación de datos, describe el mismo desplazamiento en su propia base: en 2018, Facebook aparecía en el 76,5% de las verificaciones; en 2022, WhatsApp estaba presente en el 92,4%.

Dos límites separan estas fuentes de lo que la sección 1.2 muestra para los Estados Unidos. Ninguna de ellas es un estudio sobre compartir con registro de conducta, como el de Guess et al. (2019), ni una comparación de partidos por lo que sus parlamentarios comparten, como la de Lasser et al.: lo que hay son verificaciones, lo que es curado por una agencia, y un análisis de contenido de dos emisores. Y la regla del Capítulo 2, según la cual la proporción de falsas entre las declaraciones verificadas no es la proporción entre todas las declaraciones, vale aquí íntegramente. Que la desinformación brasileña tenga el formato de la estadounidense, o lo contrario, es una hipótesis, y este capítulo no la asume.

1.7 Discurso de odio: definiciones, medidas y datos

Entre las fuentes verificadas, la expresión 'discurso de odio' tiene una definición explícita en Hangartner et al. (2021): lenguaje hostil contra un grupo o una persona por una característica innata, real o percibida; los autores seleccionaron tuits xenófobos y racistas mediante una combinación de diccionario, análisis de sentimiento y anotación manual. Los demás estudios operacionalizan de otro modo, y la Tabla 1.3 reúne lo que midió cada uno.

TABLA 1.3

Cómo miden el discurso de odio cinco estudios verificados, y lo que cada medida no permite

Estudio	Medida de "discurso de odio"	Muestra	Resultado	Lo que la medida no permite
Siegel et al. (2021)	Diccionario ampliado por aprendizaje automático y clasificación por comunidades de Reddit; el diccionario exige insulto o término despectivo	Más de 750 millones de tuits sobre la elección estadounidense de 2016 y casi 400 millones de tuits de una muestra aleatoria de usuarios estadounidenses	Sin aumento persistente en la campaña ni después de la elección	Hablar de otras plataformas; la definición es conservadora

Estudio	Medida de "discurso de odio"	Muestra	Resultado	Lo que la medida no permite
Marino-Fages y Martínez-Pozo (2026, documento de trabajo)	Clasificador BERTimbau y diccionario, por blanco (homofobia, racismo, sexismo, insultos, odio político)	Más de 37,6 millones de tuits geolocalizados en Brasil, 2017-2019	Aumento después de la elección de 2018, mayor donde Bolsonaro tuvo menos votos, en homofobia y sexismo	Medir el costo social percibido; la exactitud del clasificador es del 77% al 79%
Álvarez-Benjumea y Winter (2018)	Nota de hostilidad (7 o más) dada a comentarios de un foro experimental	180 participantes de una plataforma de crowdsourcing	Comentarios hostiles: 5,88% en la línea de base, 0,99% con remoción moderada; contradiscurso sin efecto significativo	Generalizar: participantes pagados, foro anónimo
Hangartner et al. (2021)	Tuits xenófobos y racistas, seleccionados por diccionario, sentimiento y anotación	1.350 usuarios de Twitter	El contradiscurso basado en la empatía aumentó la eliminación y redujo la creación de discurso xenófobo	Efecto sobre quien solo lee; humor y advertencia de consecuencias sin efecto consistente
Fazzano y Gallo (2015)	Homofobia como clase de respuestas (golpear, insultar, torturar, apalear, excluir, discriminar), definida por la función	Datos de Disque 100 (línea telefónica federal de derechos humanos de Brasil) de 2011: 6.809 relatos de violaciones contra personas LGBT	Registros de denuncias oficiales; no miden la prevalencia directa de la agresión	Prevalencia: son relatos, y no el fenómeno

Nota. Cifras de los artículos. Marino-Fages y Martínez-Pozo es un documento de trabajo que, según la portada, no pasó por revisión de pares. Los estudios se tratan en el Capítulo 5 (sección 5.5), en el Capítulo 10 y en el Capítulo 12.

Cuatro puntos organizan la tabla. El primero es que la definición decide la cifra: Siegel et al. dicen que su definición de diccionario es conservadora, pues exige que el tuit contenga un insulto o un término despectivo, y Marino-Fages y Martínez-Pozo clasifican por blanco y obtienen un aumento concentrado en homofobia y sexismo, sin un patrón comparable para insultos generales, odio político o blancos políticos. El segundo es la plataforma y el recorte de la muestra: el estudio estadounidense se limita a Twitter y advierte que observar cierto discurso en un momento no significa que haya aumentado, dado el carácter intermitente de la plataforma, en la que los temas cobran relevancia por poco tiempo y vuelven al equilibrio; el brasileño usa solo tuits publicados los lunes y geolocalizados en Brasil, y asume la premisa de que esos dos recortes representan el universo de la plataforma. El tercero es que "discurso" y "acto" son medidas diferentes, y ninguno de los cinco estudios mide actos

contra el grupo (la laguna se discute en la sección 5.5). El cuarto es que los resultados no se contradicen ni se confirman: miden cosas diferentes en poblaciones diferentes.

Del lado del análisis de la conducta, la literatura es escasa. Mizael (2018) encontró apenas cuatro artículos analítico-conductuales sobre homosexualidad en 13 revistas y señala como laguna los análisis funcionales de conductas prejuiciosas, como los chistes mantenidos por la comunidad verbal; y Fazzano y Gallo (2015) definen la homofobia como una clase de respuestas de topografías diversas, definida por la función, y no por la forma. Es la definición de la que parte el Capítulo 5.

1.8 Interpretaciones rivales y límites

La lectura propuesta en este capítulo es que los datos empíricos revelan magnitudes inferiores a lo que el debate público supone, y que cada cifra es función estricta del instrumento utilizado. Esta perspectiva se enfrenta a tres interpretaciones rivales. La primera sostiene que la asimetría partidaria sería un hecho consolidado y que el capítulo adoptaría un escepticismo desmedido; la prueba decisiva entre esas lecturas exigiría un diseño que mantuviera bajo control riguroso la oferta total de contenido y la exposición individual efectiva —salvedad que los propios autores de Garrett y Bond (2021) y Guess et al. (2019) formulan. La segunda tesis rival argumenta que, sin verificación profesional de datos, cualquier medición es inviable, lo que hace de las agencias el único termómetro disponible; la respuesta del Capítulo 2 es que la verificación mide lo que fue verificado, y no lo que circula. La tercera es que la definición por la intención vuelve imposible la medida y debería abandonarse; la de dominio, que usan los estudios, es una respuesta pragmática, y aquí se toma como tal.

Los límites del capítulo son los de las fuentes. La mayoría de las medidas son de los Estados Unidos y de una plataforma, en el período de una o dos elecciones; la única base brasileña de gran escala, entre las de discurso, es un documento de trabajo sin revisión de pares; y las cifras de prevalencia de información falsa son, en el mejor de los casos, de exposición y de compartir, y no de efecto. No se compara a Brasil con los Estados Unidos, y compartir una fuente no se trata como creencia ni como acto.

Síntesis

Cada definición de información falsa impone exigencias metodológicas propias (intención del emisor, origen del contenido, confiabilidad del dominio o veracidad fáctica), y las investigaciones tienden a sustituir el análisis de la declaración por la clasificación de la fuente debido a la inaccesibilidad de la intención privada (1.1). La circulación empírica de noticias falsas que se comparten individualmente resultó rara en los datos de Guess et al., con mayor incidencia entre las personas mayores, y la exposición pasiva no equivale a impacto causal (1.2). La asimetría partidaria registrada entre parlamentarios en los Estados Unidos por el criterio de los dominios compartidos parece constituir un fenómeno peculiar de aquel contexto institucional (1.3). Las relaciones entre posicionamiento ideológico y desinformación tienen carácter correlacional, dependen del conjunto de ítems seleccionados y del ambiente de oferta, y una parte sustancial de la disparidad de exactitud se altera cuando se manipulan las contingencias de reforzamiento (1.4). El autoinforme en encuestas consiste en conducta verbal bajo control de las contingencias de la recolección, y el muestreo arbitrario de estímulos puede sesgar el resultado de las medidas (1.5). En Brasil, el cuadro empírico se limita a análisis de contenido, informes de coyuntura y bases de agencias de verificación, y no hay registros conductuales directos de lo que se comparte en redes abiertas (1.6). Por último, las mediciones de discurso de odio varían radicalmente según los criterios léxicos y la plataforma de recolección adoptados, y no captan la emisión de actos concretos contra los grupos que son su blanco (1.7).

Los capítulos siguientes movilizan directamente este marco: los Capítulos 2 y 3 operan bajo el principio de que toda tasa de falsedad es función del instrumento de medida; el Capítulo 5 adopta las definiciones funcionales de la sección 1.7; el Capítulo 6 se fundamenta en la evidencia de que compartir representa una fracción ínfima de la exposición; y el Capítulo 12 examina la discrepancia entre lo que el individuo relata verbalmente y lo que efectivamente hace ante contingencias reales.

Referencias

Agência Lupa. (2026, 27 de julio). *Anatomia da desinformação eleitoral* [Informe]. 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), Artículo 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), Artículo eaau4586. https://doi.org/10.1126/sciadv.aau4586

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), Artículo 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), Artículo 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), Artículo 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), Artículo 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-researc/168076277c

instituto WALDEN4 | ACADEMIA WALDEN4

ENCICLOPÉDIA IW4

Los conceptos de este libro, una entrada a la vez

Reforzamiento, control de estímulos, equivalencia, práctica basada en la evidencia: entradas técnicas escritas con claridad, con las fuentes al pie de cada una. Para estudiar, repasar y citar.

80+ entradas abiertas

Acceda, gratis y sin registro

walden4.com.br/enciclopedia

CAPÍTULO

2

Lo que dicen las agencias de verificación (y lo que no pueden decir)

En julio de 2026, la Agência Lupa, agencia brasileña de verificación de datos, divulgó el informe *Anatomia da desinformação eleitoral*, con la base de sus verificaciones en cinco ciclos electorales, de 2016 a 2024. Entre los 23 políticos con más de 30 frases evaluadas, Jair Bolsonaro tuvo la mayor proporción de frases clasificadas como "Falso": 102 de 236, o 43,2%, en las campañas presidenciales de 2018 y 2022. Luiz Inácio Lula da Silva aparece en séptimo lugar: 22 de 83 frases, o 26,5%, en campañas electorales que disputó "recientemente" (Agência Lupa, 2026, p. 12, 19–20).

En otro frente, Aos Fatos, también agencia brasileña de verificación de datos, mantuvo, de 2019 a 2022, un contador de las declaraciones "falsas o distorsionadas" de Bolsonaro. El 22 de diciembre de 2022 informó que dejaba la presidencia con 6.675 (Ribeiro & Menezes, 2022); en julio de 2025, el mismo medio habló de 6.685 (Menezes & Faustino, 2025). Para Lula, Aos Fatos no tiene un contador equivalente: publicó, a fines de 2023 y de 2024, recopilaciones de las afirmaciones que él más repitió (Ribeiro, 2023; Menezes, 2024).

Poner 43,2% al lado de 26,5%, o 6.685 al lado de un ranking de afirmaciones repetidas, da la impresión de comparar. En este capítulo se sostiene que, con los datos publicados hoy, ninguna de las dos operaciones constituye una comparación legítima, y se explicita lo que faltaría para que lo fuera.

No es escepticismo contra las agencias. Es la lectura que ellas mismas piden: Lupa escribe, en el mismo informe, que optó por no producir rankings de quién "mintió más", porque el volumen de verificaciones sobre cada político depende de la exposición pública, de la frecuencia de entrevistas y debates, del cargo y del estilo de comunicación (Agência Lupa, 2026).

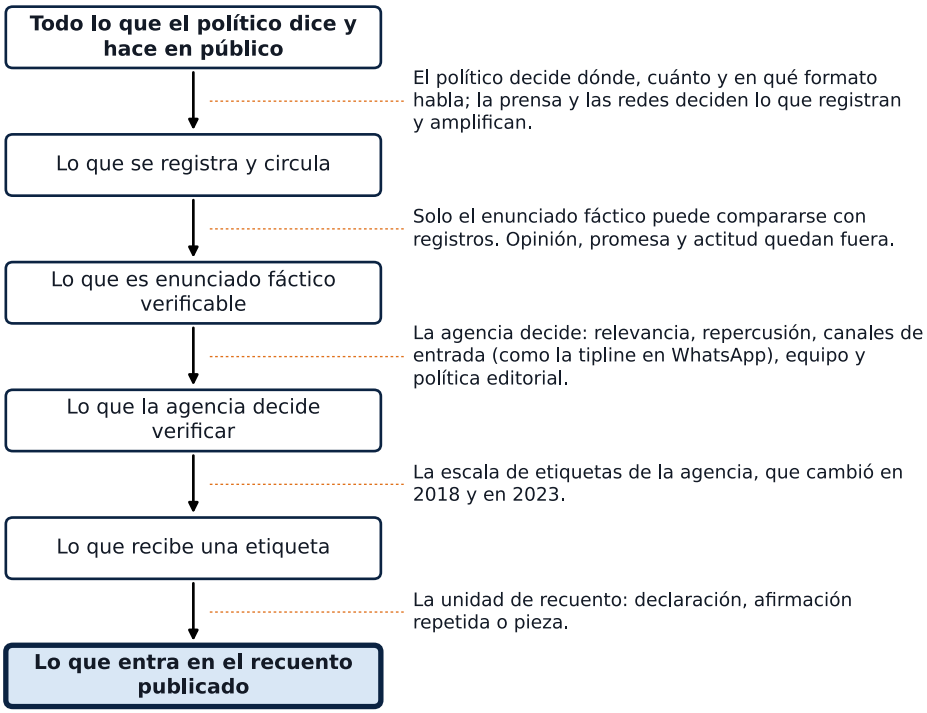
El capítulo cumple también una función en el libro: fija las reglas con que los capítulos siguientes leen datos de agencias. Cuando un dicho aparezca como "clasificado como falso por la agencia X", quiere decir exactamente eso, y nada más.

2.1 Dos series de contingencias

Un dicho de un político que llega al recuento de una agencia pasó por dos series de contingencias. La primera es la del político: dónde, para quién, con qué frecuencia y en qué formato habla. La segunda es la de la agencia: qué entra en su radar, qué considera relevante y verificable, qué escala usa para clasificar y cómo cuenta. La Figura 2.1 resume el recorrido. Cada flecha es una selección, y solo la última etapa llega al lector.

FIGURA 2.1

Del enunciado a la clasificación: dónde ocurre la selección



Solo la última etapa llega al lector: las proporciones publicadas son fracciones de lo que pasó por las anteriores.

Nota. Esquema original del autor. No representa proporciones: las etapas no tienen un tamaño conocido. Los mecanismos anotados en las flechas vienen de la propia Lupa (Agência Lupa, 2026) y, para la unidad de recuento, de las fuentes de la Tabla 2.1.

Tres puntos de la figura merecen detalle, porque las propias agencias los documentan.

Primero, lo que se verifica depende de cuánto se expone el político. Lupa observa que los candidatos que hablan más y son más incisivos producen más afirmaciones susceptibles de verificación y, por eso, pueden "tener más 'Falsos' y destacar erróneamente como el más mentiroso" (Agência Lupa, 2026, p. 19). El contador de Aos Fatos muestra el mismo efecto en la escala de un día: una única entrevista, en el Flow Podcast el 8 de agosto de 2022, rindió al menos 26 declaraciones falsas o distorsionadas en cinco horas y 21 minutos, y el récord del mandato fue el 15 de enero de 2021, con 61 dichos verificados (Santos, 2022). Los dos casos sugieren que el recuento acompaña, en parte, el tiempo de exposición.

Segundo, lo que entra en el radar de la agencia depende de sus canales de entrada. Lupa mantiene una tipline en WhatsApp e integra programas de verificación de Meta, lo que, en sus palabras, tiende a concentrar parte del flujo de contenidos recibidos en plataformas de Meta, de modo que los resultados sobre plataformas deben leerse como "un retrato del universo monitoreado por Lupa" (Agência Lupa, 2026, p. 82). Y la base del informe refleja solo lo que la agencia publicó en la sección de Elecciones, condicionado a decisiones editoriales y a criterios de relevancia periodística.

Tercero, la selección puede leerse como conducta de la propia agencia. Kniess et al. (2024) parten de un escenario en el que el periodismo enfrenta una crisis de legitimidad y de confianza y preguntan en qué medida dos agencias ligadas a empresas de comunicación tradicionales, Estadão Verifica y Fato ou Fake, buscan reforzar su propia autoridad epistémica, esto es, la credibilidad como fuente de lo que es verdadero, por medio de la verificación. Para responder, recogieron todos los textos que las dos agencias publicaron con el nombre de Bolsonaro en el título entre enero y mayo de 2022 y con el de Lula entre enero y mayo de 2023, 119 verificaciones en total (95 de Estadão Verifica y 24 de Fato ou Fake), y sometieron ese corpus al análisis de contenido. Las agencias concentraron la cobertura en asuntos semejantes en los dos gobiernos, y la investigación dio énfasis a fuentes del propio periodismo, esto es, a materiales propios o de otras empresas de comunicación tradicionales. Según los autores, eso sugiere límites a la transparencia de la investigación, porque muchas de las informaciones usadas ya pasaron por filtros editoriales no siempre claros para el público.

En lenguaje analítico-conductual, la verificación es conducta verbal del verificador. Está bajo control de estímulos antecedentes (el enunciado del político y la audiencia), pero también de contingencias de reforzamiento institucionales (como la política editorial) y de consecuencias que van más allá del enunciado verificado. Eso no la vuelve arbitraria. Vuelve el recuento dependiente de un

proceso de selección que necesita describirse para poder leerse. Que entre esas consecuencias esté, por ejemplo, la credibilidad ante la audiencia es una hipótesis compatible con Kniess et al. (2024), no un dato de este capítulo.

2.2 Lo que cuenta cada fuente

Antes de comparar números, pregunte de cada uno qué se contó. El Recuadro 2.1 reúne las cinco preguntas y la Tabla 2.1 las responde para las fuentes usadas en este capítulo.

RECUADRO 2.1

Cinco preguntas antes de comparar recuentos

1. ¿Cuál es la *unidad*: frase, declaración, afirmación repetida, pieza o mención?
2. ¿Cuál es el *denominador*, esto es, lo que se verificó, y se publica?
3. ¿Cuál es el *período* y la *situación*: campaña, mandato, año civil?
4. ¿Cuál es la *escala* de clasificación, y cambió?
5. *¿Quién eligió* lo que entró, y con qué regla?

TABLA 2.1

Lo que cuenta cada fuente y lo que publica como denominador

Fuente	Unidad contada	Lo que se publicó	Denominador publicado	Período y situación
Lupa, informe (Agência Lupa, 2026)	Frase verificada (unidad de análisis dentro de una verificación)	Frases “Falso” por político, entre los que tienen más de 30 frases	Sí, por político: 236 (Bolsonaro) y 83 (Lula)	Campañas electorales de 2016 a 2024; Bolsonaro, 2018 y 2022; Lula, ciclos no discriminados
Aos Fatos, contador de Bolsonaro (Santos, 2022; Ribeiro & Menezes, 2022; Menezes & Faustino, 2025)	Declaración “falsa o distorsionada” (“imprecisa” e “incorrecta”, en 2025)	6.042 (ago./2022); 6.675 (dic./2022); 6.685 (jul./2025)	No	Mandato, ene./2019 a dic./2022

Fuente	Unidad contada	Lo que se publicó	Denominador publicado	Período y situación
Aos Fatos, recopilaciones de Lula (Ribeiro, 2023; Menezes, 2024)	Afirmación repetida	Ranking de las más repetidas, con el número de ocasiones (21 veces en 2023; al menos 24 en 2024)	No	Año civil de 2023 y de 2024; mandato en curso
Aos Fatos, balance de diez años (Menezes & Faustino, 2025)	Verificaciones publicadas; menciones en piezas de desinformación	Más de 19 mil verificaciones; cerca de 15 mil declaraciones de políticos y autoridades; menciones: Lula, 677, y Bolsonaro, 566	Total general, sin recorte por político ni por mandato	2015 a 2025
Kniess et al. (2024)	Verificación (pieza) con el nombre del presidente en el título	119 verificaciones: 95 de Estadão Verifica y 24 de Fato ou Fake	Sí, del corpus	Ene.–may./ 2022 (Bolsonaro) y ene.–may./ 2023 (Lula)
Recuero et al. (2025)	Verificación (pieza)	811 verificaciones de tres agencias	Sí, del corpus	Elecciones de 2018 y 2022

Nota. Todas las fuentes son verificaciones publicadas por las propias agencias o estudios sobre ellas; en todas, quien elige lo que entra es la agencia (pregunta 5).

La tabla hace evidente lo que el debate público suele ocultar. Solo Lupa publica numerador y denominador individualizados por político — restringidos, sin embargo, a períodos de campaña, y no a mandatos. Aos Fatos, por su parte, divulga para Bolsonaro un numerador desprovisto de denominador; y, para Lula, informa afirmaciones repetidas, que constituyen una unidad de análisis distinta. Sus totales generales (más de 19 mil verificaciones; cerca de 15 mil declaraciones de políticos y autoridades y casi 4.000 piezas de desinformación desde 2015) no vienen recortados por político ni por mandato (Menezes & Faustino, 2025) y no sirven de denominador para ninguno de los dos.

2.3 Por qué la proporción entre las verificadas no es la proporción entre las dichas

La fracción publicada por una agencia es una proporción entre enunciados que pasaron por la selección de la Figura 2.1. Sea p la proporción de enunciados fácticos falsos en lo que el político dice; sF, la probabilidad de que un enunciado falso sea verificado; y sV, la de que lo sea uno verdadero. La proporción observada entre las verificadas es

$$p{\cdot}s\mathrm{F} \div (p{\cdot}s\mathrm{F} + (1 - p){\cdot}s\mathrm{V}),$$

que solo coincide con p cuando sF = sV — o sea, en la hipótesis de que la agencia verifique enunciados falsos y verdaderos con idéntica probabilidad. Ninguno de los informes publicados autoriza esa premisa de igualdad, ni permite estimar la razón de selección $r = s\text{F} \div s\text{V}$. Lupa informa que verifica también para confirmar lo que es verdadero ("una función central (y frecuente) del fact-checking") y que el 34,4% de las frases verificadas recibió "Verdadero" (Verdadeiro) (Agência Lupa, 2026, p. 12, 29), lo que muestra que la selección no es solo de sospechas, pero no dice con qué probabilidad relativa.

La Tabla 2.2 muestra, con números hipotéticos, lo que la razón r hace con la proporción observada.

TABLA 2.2

Ilustración hipotética: la misma fórmula con razones de selección diferentes

Tasa real de falsedad (p)	Razón de selección (r = sF ÷ sV)	Proporción observada entre las verificadas
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, elegidos para ilustrar la fórmula; no son datos ni estimaciones sobre ningún político. La proporción observada es p·r ÷ (p·r + 1 – p).

La última fila es el punto. Un político con la mitad de la tasa real de falsedad (10% contra 20%) aparece con una proporción observada mayor (30,8%) que otro con una tasa real de 20% y razón 1,5 (27,3%), si la selección favorece más la verificación de lo que es falso en el primero. El mismo razonamiento se aplica a discrepancias entre agencias distintas o entre diferentes intervalos temporales. La conclusión no es la invalidez de esos datos, sino que la diferencia entre proporciones observadas combina dos procesos independientes — el repertorio emitido por el político y el filtro de selección de la agencia —, variables que las bases publicadas no permiten aislar.

Hay además una segunda fuente de imprecisión. Las frases de un mismo debate, entrevista o sesión de preguntas no son independientes: comparten tema, audiencia y momento. Lupa recuerda que, en las verificaciones de debates y entrevistas, un único reportaje puede analizar decenas de frases de candidatos (Agência Lupa, 2026). El caso del Flow Podcast (sección 2.1) es lo mismo visto desde el lado de Aos Fatos.

2.4 Lo que Lupa permite decir

Lupa es la única fuente examinada que publica numerador y denominador por político, con la misma escala para todos. Es, por eso, el material más cercano a una comparación. La Tabla 2.3 reúne lo que el informe dice sobre los dos presidentes, y la Figura 2.2 los sitúa entre los diez políticos del ranking publicado.

TABLA 2.3

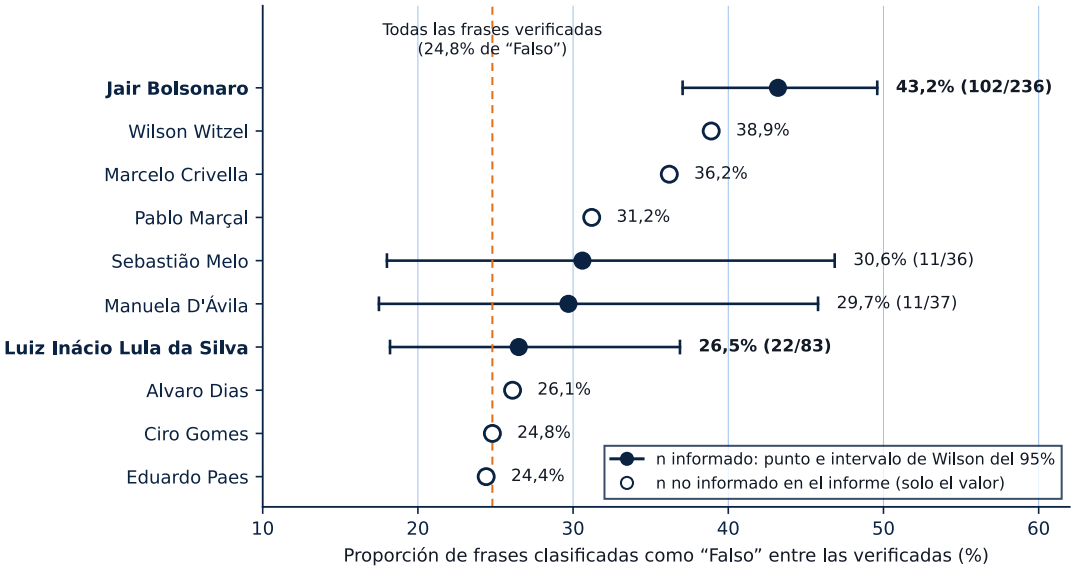
Lo que el informe de Lupa dice sobre Bolsonaro y Lula

	Jair Bolsonaro	**Luiz Inácio Lula da Silva**
Ciclos considerados	Campañas presidenciales de 2018 y 2022	"Campañas electorales que disputó recientemente"; ciclos no discriminados
Frases verificadas	236	83
Frases "Falso"	102	22
Proporción de "Falso"	43,2%	26,5%
Posición entre los 23 políticos con más de 30 frases	1.º	7.º
Intervalo de Wilson del 95%, calculado por el autor	37,1% a 49,6%	18,2% a 36,9%
Patrón de las falsedades, según el análisis lingüístico de Lupa	Números incorrectos; afirmaciones absolutas y engañosas; "nunca", "jamás", "ninguno", "siempre"	Reivindicación de autoría o de carácter pionero de políticas públicas

Nota. Fuente: Agência Lupa (2026). El intervalo de Wilson delimita el margen de confianza del 95% para una proporción bajo la premisa teórica de que las frases verificadas representaran una muestra aleatoria independiente del repertorio emitido por el hablante. Como los ítems verificados comparten contextos comunes y resultan de filtros editoriales deliberados (sección 2.3), esa condición de independencia no se satisface estrictamente; el intervalo opera aquí solo como estimador del orden de magnitud de la imprecisión derivada del tamaño de la muestra, y no como prueba de hipótesis confirmatoria.

FIGURA 2.2

Proporción de frases clasificadas como "Falso" entre las verificadas por Lupa, por político



Nota. Valores del gráfico de la p. 20 del informe de Lupa, que considera solo actores con más de 30 frases verificadas (Agência Lupa, 2026); el número de frases viene del texto y solo está disponible para cuatro políticos. El orden es el del informe y no mide "quién miente más" (sección 2.3). Los políticos son de partidos, cargos y esferas diferentes. Intervalos de Wilson del 95% calculados por el autor.

Cuatro observaciones impiden que la Tabla 2.3 se convierta en una conclusión apresurada.

Es frase, no dicho ni pieza. La unidad del informe es la frase verificada dentro de un chequeo (checagem). La base tiene 632 publicaciones, 397 chequeos y 235 verificaciones (verificações), que generaron 3.366 unidades de análisis, 3.080 de chequeos y 286 de verificaciones (Agência Lupa, 2026).

Es campaña, no mandato. Los 236 y los 83 vienen de campañas electorales. Para Bolsonaro, los ciclos son los de 2018 y 2022, en los que fue candidato y, en el segundo, presidente candidato a la reelección. Para Lula, el informe habla de campañas que disputó "recientemente" y no las discrimina. La situación del hablante (candidato, presidente, oposición) cambia lo que se dice y lo que se verifica, y la tabla no la controla.

"Falso" es una etiqueta entre varias. En el conjunto de la base, "Falso" suma 765 registros (24,8%); "Exagerado", 502 (16,3%); y "Verdadero, pero... / Falta contexto" (Verdadeiro, Mas... / Falta Contexto), 434 (14,1%) (Agência Lupa, 2026). Los dos últimos, juntos, suman 30,4%, más que "Falso", y Lupa observa que, en cuatro de los cinco ciclos, superaron a "Falso" (p. 27). El informe publicado

no discrimina la distribución de las demás etiquetas individualizada por político. Una comparación hecha solo con "Falso" deja fuera la mayor parte de lo que la agencia clasifica como problemático en el conjunto.

La escala cambió. Lupa empezó con ocho etiquetas; incorporó "Subestimado" en 2018; y, en 2023, "En la mira" (De olho), "Aún es pronto" (Ainda é cedo) y "Verdadero, pero..." dieron lugar a "Falta contexto" (Agência Lupa, 2026, p. 28). Para los análisis, los propios investigadores de Lupa fusionaron "Verdadero, pero..." y "Falta contexto". Cualquier serie que atraviese 2023 necesita hacer lo mismo.

DATO	En la base de Lupa, el 43,2% de las 236 frases verificadas de Bolsonaro en las campañas presidenciales de 2018 y 2022 se clasificó como "Falso", frente al 26,5% de las 83 frases verificadas de Lula en campañas que disputó recientemente. La proporción de "Falso" en la base entera es del 24,8%.
INTERPRETACIÓN	Entre las frases que Lupa eligió verificar en campañas, la proporción de "Falso" fue mayor para Bolsonaro que para Lula y que la media de la base. Los intervalos de Wilson (37,1% a 49,6% y 18,2% a 36,9%) casi se tocan; indican una imprecisión considerable, sobre todo para Lula.
HIPÓTESIS	Parte de la diferencia puede reflejar estilos de habla (Lupa describe números y afirmaciones absolutas en un caso y autoría y carácter pionero en el otro), la situación del hablante en cada ciclo y la razón de selección *r* de la sección 2.3. Ninguna de las tres se puso a prueba aquí.

2.5 Lo que Aos Fatos permite decir

Aos Fatos publica dos tipos de recuento sobre los presidentes, y cada uno tiene unidad y criterio propios: un contador continuo, mantenido entre 2019 y 2022, y recopilaciones anuales sobre Lula. Los dos apartados siguientes dicen lo que cada uno permite concluir.

El contador. Entre 2019 y 2022, Aos Fatos mantuvo un contador de las afirmaciones falsas del presidente, que monitoreaba "entrevistas, transmisiones en vivo, discursos y publicaciones desde la toma de posesión, en enero de 2019" (Santos, 2022). El 17 de agosto de 2022 marcaba al menos 6.042; el 22 de diciembre, al final del mandato, 6.675, "un promedio de 4,6 por día" (Ribeiro & Menezes, 2022); en julio de 2025, el texto del mismo medio habla de 6.685, unas veces como declaraciones "falsas o imprecisas", otras como "incorrec-

tas" (Menezes & Faustino, 2025). En los textos publicados por el medio, no hay una justificación explícita para la discrepancia de diez registros ni la especificación del criterio adoptado para la inclusión de la etiqueta 'imprecisa'.

Dos cosas, sin embargo, están claras. La primera es que el contador publica un numerador sin denominador: no figura, en los textos divulgados por el medio, el total de declaraciones del presidente que se verificaron, de modo que 6.675 no expresa la proporción de ningún conjunto. La segunda es que la unidad, la declaración, es acumulativa y acompaña la exposición, como mostró el caso del Flow Podcast. El promedio de 4,6 por día es coherente con 1.461 días de mandato (6.675 ÷ 1.461 = 4,57, cálculo del autor), pero es un promedio por día de mandato, y no por dicho ni por hora de exposición.

Las recopilaciones de Lula. En 2023, la recopilación se hizo "con base en los discursos, entrevistas y transmisiones en vivo del presidente" (Ribeiro, 2023). En 2024, Aos Fatos "analizó todos los discursos, pronunciamientos y entrevistas" de Lula y separó las declaraciones susceptibles de verificación; en el transcurso del año, verificó las más repetidas, y, a fines de diciembre, las clasificadas como "Falso" o "No es tan así" (Não é bem assim) se contaron para formar un ranking (Menezes, 2024).

La unidad, por lo tanto, es la afirmación repetida, y el ranking responde a una pregunta diferente de la del contador: "¿qué falsedades repitió más?", y no "¿cuántas declaraciones falsas hizo?". En 2023, la más repetida fue la de que había heredado 14 mil obras paralizadas (21 veces; clasificada como falsa) (Ribeiro, 2023). En 2024, fue la de que el gobierno del PT (Partido dos Trabalhadores) acabó con el hambre en 2014 (al menos 24 ocasiones; clasificada como falsa), y otra, repetida en al menos 17 ocasiones, fue la de que sería el único presidente de la República sin título universitario (clasificada como falsa) (Menezes, 2024). Para Bolsonaro, el balance de diez años informa que la afirmación más repetida fue la de que no hubo ninguna denuncia de corrupción en su gobierno (Menezes & Faustino, 2025).

Los dos conjuntos solo se volverían comparables si los dos presidentes fueran sometidos a las dos preguntas: el ranking de afirmaciones repetidas y el total de declaraciones verificadas, en el mismo período y con el mismo criterio.

DATO	Aos Fatos publicó 6.675 declaraciones "falsas o distorsionadas" de Bolsonaro en el mandato (dic./2022) y 6.685 en 2025; para Lula, publicó rankings anuales de afirmaciones repetidas y clasificadas como falsas. Ninguno de los dos trae el total de declaraciones verificadas.

INTERPRETACIÓN	Son descripciones fieles del recuento de cada serie, pero de unidades diferentes (declaración y afirmación repetida) y sin denominador. No permiten decir quién "dijo más falsedades" ni quién las dijo en mayor proporción.
HIPÓTESIS	Que el contador de Bolsonaro haya crecido a un ritmo parecido al de la exposición (entrevistas largas, transmisiones en vivo) es plausible por el caso del Flow Podcast, pero no se ha puesto a prueba.

2.6 Dichos y menciones: de quién es la falsedad

Un error de unidad frecuente es tratar como equivalentes dos tipos de pieza: el chequeo de lo que el político dijo y la verificación de un contenido que circula sobre él. Lupa los separa. El "chequeo" (checagem) es la verificación de dichos, declaraciones o afirmaciones específicas, generalmente atribuidas a personas públicas; la "verificación" (verificação), la de contenidos virales o potencialmente virales, sin autoría identificada (Agência Lupa, 2026).

Aos Fatos informa, en su balance de diez años, que Lula y Bolsonaro fueron las figuras públicas más verificadas y también las más citadas en piezas de desinformación: 677 menciones para Lula y 566 para Bolsonaro. Y añade que, "mientras el petista es más atacado por las piezas engañosas, Bolsonaro, en su mayor parte, ve sus acciones infladas" (Menezes & Faustino, 2025). Kniess et al. (2024) constataron que más de la mitad de las 119 verificaciones que componían su corpus se destinaba a refutar rumores desfavorables a Lula.

Esos números describen lo que circula *sobre* cada político, no lo que cada uno dijo. Contar piezas de desinformación que citan a un político como si fueran dichos suyos es un error de unidad. En los capítulos siguientes, se trabaja solo con dichos atribuidos al propio político, y el contenido sobre él se trata como otro objeto.

2.7 ¿Concuerdan las agencias entre sí?

Una verificación es una clasificación, y las clasificaciones pueden divergir. Hay tres fuentes de variación.

La primera es la escala: la de Lupa cambió dos veces (sección 2.4). La segunda es el contenido: Recuero et al. (2025), en 811 verificaciones de tres agencias brasileñas en las elecciones de 2018 y 2022, encontraron una transición de contenidos completamente falsos a engañosos y de contenidos sin fuentes a contenidos que usan fuentes reales. Lo que se clasifica cambió, y con ello la

distribución de las etiquetas. La tercera es la concordancia entre las propias agencias: hay estudios empíricos sobre la consistencia entre verificadores en debates electorales (p. ej., Lelo, 2023), pero cuyos datos puntuales no se incorporan a este cómputo (véase el Apéndice D).

Mientras la tercera pregunta no pueda responderse con datos, cada clasificación se trata como atribuida a la agencia que la hizo, y no como un hecho establecido.

2.8 Lo que la verificación no alcanza

La verificación opera sobre enunciados fácticos, que se comparan con registros: datos oficiales, documentos, grabaciones. El Análisis de la Conducta ofrece una distinción útil para saber dónde termina eso.

CONCEPTO EN USO

Creencia y actitud como conducta verbal (Guerin, 1994)

Guerin trata las actitudes y las creencias como conducta verbal moldeada por comunidades verbales, y no como relatos de un estado interno. Una *creencia* ("los reactores nucleares son inseguros") puede impugnarse con evidencia o con el enunciado contrario. La *actitud* correspondiente ("no me gusta el uso de la energía nuclear") no puede atacarse del mismo modo: el hablante puede responder, en la formulación de Guerin, que, "a pesar de todo lo que has dicho", sigue sin gustarle (traducción mía). Por eso, presentar un tema como actitud, y no como creencia, es reforzado con frecuencia, y la diferencia entre las dos formas es una contingencia de la comunidad de oyentes, no una propiedad del enunciado (Guerin, 1994).

CONCEPTO EN USO

Autoclítico (Guerin, 1994)

Los autoclíticos son calificaciones añadidas al enunciado que modifican su efecto sobre el oyente, como "me parece que...", "siempre había creído que..." o "me gustan un poco los perros" dicho a una sociedad de amantes de los gatos (Guerin, 1994; término de Skinner, 1957, citado por Guerin). Pueden reducir consecuencias negativas y permiten recalificar lo que se dijo después de una réplica negativa.

Lo que las agencias evalúan se restringe a enunciados fácticos — equivalentes, en el vocabulario analítico-conductual de Guerin (1994), a la clase de las creencias, que mantienen correspondencia con estímulos antecedentes verificables. En cambio, las declaraciones formuladas como actitud (por ejemplo, 'este gobierno es el peor de la historia') o mediadas por calificadores autoclíti-

cos ('a mi modo de ver', 'solo estoy preguntando') se encuentran predominantemente bajo control de consecuencias sociales inmediatas de la audiencia, y se sitúan fuera del ámbito de aplicación de las etiquetas de veracidad fáctica. Además, Guerin (1994) demuestra que los enunciados en forma de actitud operan con frecuencia como mandos: la supervisora que declara en público que no le gustan los empleados que toman café durante la jornada ejerce funcionalmente control sobre la conducta del oyente para que deje de hacerlo, independientemente de cualquier intención privada.

La consecuencia para la lectura de las agencias es directa. La etiqueta mide la correspondencia entre lo que se dijo y los registros. No mide la función de lo que se dijo en la audiencia. Dos enunciados, uno verificable y otro una actitud o una calificación autoclítica, pueden tener efectos semejantes sobre quien escucha, y solo el primero entra en el recuento. Los Capítulos 3 y 5 analizan la función; este capítulo se limita a registrar que la parte del discurso que migra de la forma fáctica a la de opinión sale del alcance de la verificación sin que la función cambie.

DATO	Guerin (1994) describe que las creencias pueden impugnarse con evidencia y las actitudes no, y que la presentación en forma de actitud es reforzada con frecuencia. Lupa y Aos Fatos verifican enunciados fácticos.
INTERPRETACIÓN	Un enunciado que pase de la forma fáctica a la de opinión deja de ser verificable, aunque lo que hace con el oyente sea parecido.
HIPÓTESIS	Que parte de esa migración sea una respuesta a las propias contingencias de la verificación (evitación del contracontrol, Capítulo 6) es una hipótesis que debe ponerse a prueba con dichos fechados, antes y después de verificaciones; no hay datos sobre eso en este capítulo.

2.9 Protocolo para una muestra comparable

Para comparar a los dos presidentes sin que la elección de los casos dependa del resultado, la regla siguiente se fijó antes de cualquier recolección.

1. **Agencia.** Una sola, Lupa, que tiene escala única y publica denominador. Mezclar agencias mezclaría escalas y criterios de selección.

2. **Ventanas.** Los primeros 45 meses de cada mandato: del 1 de enero de 2019 al 30 de septiembre de 2022 (Bolsonaro) y del 1 de enero de 2023 al 30 de septiembre de 2026 (Lula), por fecha de publicación de la verificación. Las dos tienen la misma duración y terminan antes de la primera vuelta (2 de octubre de 2022 y 4 de octubre de 2026).
3. **Unidad.** La frase atribuida al propio político en una verificación. Quedan fuera los contenidos virales sobre él y los dichos de terceros. Cuando una pieza trae varias frases, se cuenta cada una; cuando el marcado no las separa, se cuenta la pieza y se señala.
4. **Escala.** Se mantienen las etiquetas originales. "Verdadero, pero..." (antes de 2023) y "Falta contexto" (después) se agrupan, como hace Lupa; "En la mira" y "Aún es pronto" quedan aparte, como "sin veredicto".
5. **Salidas.** Número de frases por etiqueta y proporción, con intervalo de Wilson, para cada presidente; tabla separada por piezas de frase única y de varias frases.
6. **Sensibilidad.** Repetir excluyendo los períodos de campaña oficial, excluyendo debates y entrevistas largas, y por trimestre, para ver si la diferencia aparece o desaparece.
7. **Validación.** Comparar los recuentos de las campañas de 2018 y 2022 con los 102 de 236 del informe, y los de Lula con los 22 de 83, para estimar la cobertura de la recolección.
8. **Límites declarados desde ya.** La selección es de la agencia; no hay denominador de dichos totales; la exposición de los dos varía a lo largo del mandato; la recolección depende de la cobertura de la fuente.

La recolección no se hizo, por tres razones. (a) Los archivos robots.txt de Lupa y de Aos Fatos restringen la recolección automatizada por herramientas computacionales, parámetros que se respetaron rigurosamente. (b) El repositorio público ClaimReview de Google Data Commons, que agrega verificaciones estructuradas con ese metadato, presenta una cobertura histórica irregular en relación con Lupa: registra 3.369 verificaciones concentradas entre 2022 y septiembre de 2024 (con solo 12 entradas en 2019, 3 en 2020 y 4 en 2021, sin registros bajo el dominio actual), lo que hace inviable delimitar ventanas temporales simétricas. (c) El informe de Lupa no discrimina los ciclos de Lula. Los caminos abiertos son: que Lupa ceda la base del Observatorio (la base que sustenta el informe existe), o que la recolección la haga el autor, bajo su responsabilidad.

Síntesis

Una fracción publicada por una agencia es la proporción entre los enunciados que pasaron por dos series de selección, la del político y la de la agencia (Figura 2.1). Por eso, antes de comparar números, conviene preguntar de cada uno cuál es la unidad, cuál es el denominador, cuál es el período y la situación, cuál es la escala y quién seleccionó lo que entró (Recuadro 2.1).

Hechas esas preguntas, los datos publicados hoy no sostienen la comparación entre Bolsonaro y Lula. Lupa publica numerador y denominador por político (Bolsonaro, 102 de 236, o 43,2%; Lula, 22 de 83, o 26,5%), pero son fracciones de campañas, con ciclos diferentes y no discriminados, y la etiqueta "Falso" es una entre varias. Aos Fatos publica, para Bolsonaro, un numerador sin denominador (6.675 en 2022; 6.685 en 2025) y, para Lula, rankings de afirmaciones repetidas: son unidades diferentes. Y las piezas sobre un político, las menciones, no son dichos suyos.

La verificación, por último, incide estrictamente sobre enunciados fácticos (creencias, en la conceptualización de Guerin, 1994). Las formulaciones de actitud y las modulaciones autoclíticas permanecen inmunes a la verificación, aunque puedan ejercer funciones conductuales análogas sobre la audiencia. Una comparación legítima entre Bolsonaro y Lula exigiría el escrutinio de una misma agencia, delimitado por ventanas temporales simétricas de ejercicio del mandato y acompañado del respectivo denominador. Como esa base probatoria no existe en las publicaciones actuales, la regla metodológica para su eventual construcción se formula en la sección 2.9.

Referencias

Agência Lupa. (2026, 27 de julio). *Anatomia da desinformação eleitoral* [Informe]. 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*, Artículo e024020. 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 diciembre). *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 julio). *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 diciembre). *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 diciembre). *Uma retrospectiva das mentiras do governo Bolsonaro* [Reportaje en cómic]. 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.

instituto WALDEN4
walden4.com.br



PARTE II

Análisis operante de los discursos

3 Declaraciones falsas como operantes verbales

4 Actitudes políticas y prejuicio como clases de equivalencia

5 Discurso de odio: mando disfrazado y control aversivo

6 Burbujas digitales como comunidades verbales y la evitación del contracontrol

7 El voto como conducta verbal

CAPÍTULO

3

Declaraciones falsas como operantes verbales

El 28 de agosto de 2022, en un debate entre candidatos a la Presidencia transmitido por Band, Folha, TV Cultura y UOL, Luiz Inácio Lula da Silva dijo: "La menor deforestación de la Amazonía se hizo en mi gobierno". La Agência Lupa, agencia brasileña de verificación de datos, clasificó la frase como "Falso": la menor tasa de deforestación de la Amazonía Legal registrada en el sistema Prodes, del Inpe (Instituto Nacional de Pesquisas Espaciais), fue la de 2012, en el gobierno de Dilma Rousseff, con 4,6 mil km^2, la menor de la serie iniciada en 1989 (Agência Lupa, 2026). En el mismo debate, Lula dijo estar orgulloso de haber nombrado "al primer negro para la Corte Suprema"; Lupa también la clasificó como "Falso", porque antes de Joaquim Barbosa, nombrado en 2003, otros dos ministros negros habían integrado el Supremo Tribunal Federal (STF, la corte suprema de Brasil), Pedro Lessa y Hermenegildo Rodrigues de Barros (Agência Lupa, 2026). El 29 de septiembre de 2022, en un debate de TV Globo, Jair Bolsonaro dijo: "Ningún país del mundo compró vacunas en 2020"; "Falso", según Lupa, que cita un relevamiento del Poder360, medio periodístico brasileño, según el cual al menos 50 países empezaron a vacunar en 2020. Y el 3 de agosto de 2018, en una entrevista al programa Central das Eleições, de GloboNews, dijo: "Yo nunca fui homofóbico"; "Falso", en la verificación de Lupa, que recuerda la condena por daño moral colectivo confirmada por el Tribunal de Justicia de Río de Janeiro en noviembre de 2017, por declaraciones que la agencia describe como de contenido homofóbico.

En las cuatro frases hay un rasgo que no es la falsedad. Todas tienen la forma gramatical de quien constata un hecho, y todas traen uno de los términos que la propia Lupa enumera entre los que hacen absoluta una afirmación: "menor", "primero", "ninguno", "nunca" (p. 41). La pregunta que suele venir enseguida es si quien las dijo mentía, si creía en ellas o si se equivocó. Este capítulo cambia la pregunta. En lugar de inferir lo que el hablante tenía en mente,

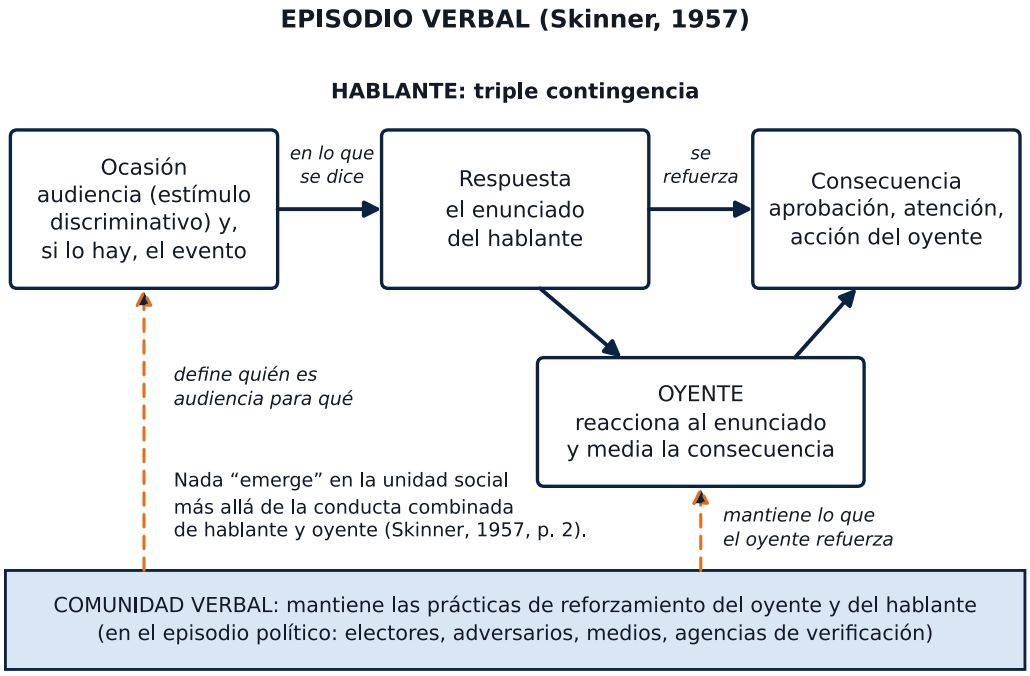
pregunta qué variables controlan la emisión de una frase así y qué esconde su forma gramatical. Es el capítulo núcleo del libro: presenta la ficha de análisis que usan los capítulos siguientes y la aplica a estos cuatro casos. No fueron elegidos al azar ni son una muestra: son los ejemplos que la propia Lupa reprodujo en el informe para ilustrar el estilo de cada político (Agência Lupa, 2026), y entran aquí porque están documentados en una fuente leída, con fecha, contexto y la clasificación atribuida a quien la hizo.

3.1 El discurso del político como conducta verbal

Skinner (1957) define la conducta verbal como un tipo de conducta necesariamente reforzada por la mediación de otras personas. La definición tiene consecuencias prácticas. Lo que un político dice no cambia directamente el mundo: cambia lo que hacen los oyentes, y es por medio de lo que ellos hacen como la conducta del hablante tiene consecuencias. Skinner llama episodio verbal total al conjunto de las conductas del hablante y del oyente, y añade que "no hay nada en un episodio así que sea más que la conducta combinada de dos o más individuos. Nada 'emerge' en la unidad social" (p. 2). La Figura 3.1 reúne el esquema.

FIGURA 3.1

El episodio verbal: la triple contingencia del hablante, mediada por el oyente y mantenida por la comunidad verbal



Nota. Esquema original del autor, a partir de Skinner (1957). No representa datos. En la ocasión, la audiencia entra como estímulo discriminativo.

Tres consecuencias de esta definición orientan el resto del capítulo. La primera es que la forma gramatical no identifica el operante. Skinner señala que términos como "declarar" o "afirmar" aplicados al tacto se refieren a procesos adicionales, los autoclíticos, y que el tacto, tomado en sí mismo, no es por ello una aserción (Skinner, 1957). Una frase en indicativo, con sujeto y predicado, es solo una frase en indicativo, y solo la historia y las variables presentes dicen si fue controlada por un evento, por otras palabras o por lo que el oyente haría después. La segunda es que la declaración de un político nunca tiene una sola audiencia. Skinner trata de las audiencias múltiples y afirma que una misma respuesta verbal puede producir efectos distintos según la audiencia en cuya presencia se emite. Cuando cada audiencia se presenta por separado, el control sobre la conducta verbal es mucho más preciso. Por otro lado, cuando hay audiencias distintas a la vez, es común que se produzcan simultáneamente consecuencias conflictivas. En un debate televisado están los electores a los que el candidato quiere convencer, los que ya lo apoyan, los adversarios en el escenario y, después, quien verifica. Estos diferentes oyentes se constituyen en audiencias conflictivas.

CONCEPTO EN USO

Triple contingencia, función × topografía y episodio verbal (Skinner, 1957)

Un operante verbal se define por lo que controla la respuesta, y no por su forma. Para Skinner (1957), ninguna respuesta puede clasificarse como un mando solo por su forma y, en general, para identificar cualquier operante verbal es preciso conocer el tipo de variables de las que la respuesta es función. La misma topografía de respuesta, es decir, la forma (p. ej., "la vacuna es peligrosa") puede, por lo tanto, ser un tacto, un intraverbal o un mando, según la ocasión, la historia y la consecuencia. En otras palabras, lo que la clasifica son las variables de control de la respuesta verbal "la vacuna es peligrosa". La *triple contingencia* es la unidad: ocasión (estímulo discriminativo), respuesta, consecuencia. En la conducta verbal, la ocasión incluye la audiencia, que es un estímulo discriminativo en cuya presencia la conducta verbal es característicamente reforzada, y la consecuencia viene del oyente. El *episodio verbal* incluye las interacciones entre hablantes y oyentes cuyas respuestas verbales ejercen múltiples funciones de estímulo sobre la conducta de unos y otros.

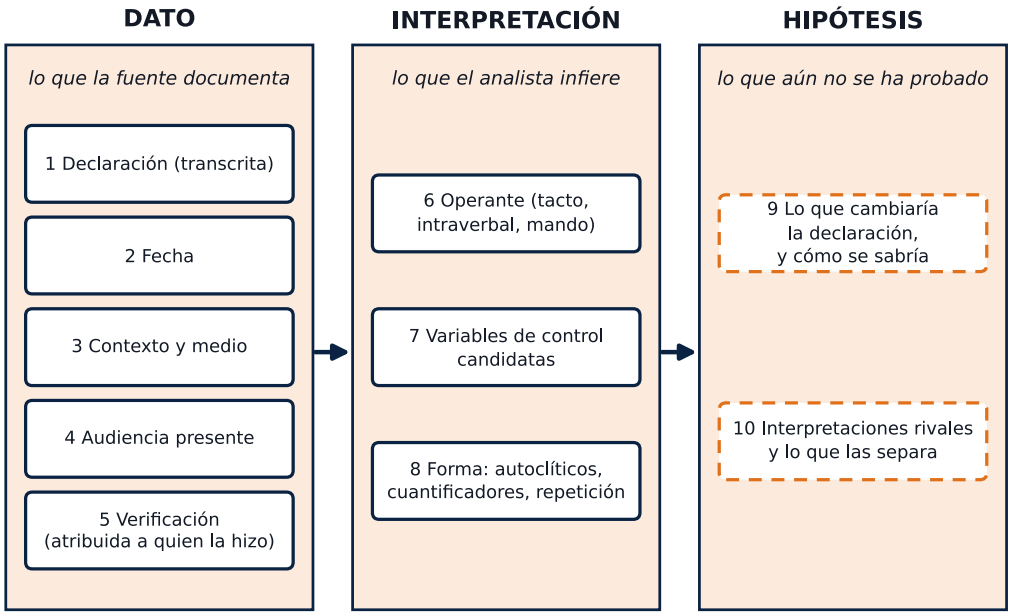
3.2 La ficha de análisis

Todo análisis de este libro pasa por la misma ficha, y existe para obligar a separar lo que se sabe de lo que se infiere. La Figura 3.2 muestra los campos. Los cinco primeros son dato: la declaración, transcrita como la trae la fuente; la fecha; el contexto y el medio; la audiencia presente, cuando la fuente la do-

cumenta; y la verificación, siempre con el nombre de quien clasificó, porque "falso" no es una propiedad de la frase, es la decisión de una agencia con un método (Capítulo 2). Los tres siguientes son interpretación: el operante probable, las posibles variables de control candidatas y los elementos verbales adicionales (autoclíticos, cuantificadores, repetición). Los dos últimos son hipótesis: lo que cambiaría la declaración y cómo se sabría, y las interpretaciones rivales con lo que las separa. Donde la fuente no documenta, la ficha escribe "no documentado" y no llena el campo por inferencia.

FIGURA 3.2

La ficha de análisis: lo que es dato, lo que es interpretación y lo que es hipótesis



En cada campo vacío, la ficha escribe "no documentado"; no lo llena por inferencia.

Contorno discontinuo: lo que, en los casos del capítulo, es hipótesis y no resultado.

Nota. Esquema original del autor. La versión para uso, con los campos en blanco, está en el Apéndice A.

Una regla de lectura acompaña la ficha y es del propio Skinner. Ante registros de la conducta de un hablante cuyas condiciones de emisión se desconocen, es posible reconstruir una historia verbal probable, pero solo raramente las inferencias sobre las variables relevantes pueden comprobarse directamente. El criterio de que se entiende un análisis causal de la conducta verbal es la capacidad de predecir la aparición de instancias y, eventualmente, de producirlas o controlarlas. Los casos de este capítulo son registros de declaraciones emitidas por terceros, sin acceso a las condiciones en que fueron emitidas;

según ese criterio, sus análisis son meras interpretaciones. Es lo que vale llamar interpretación: una lectura plausible, que dice qué variables podrían controlar la respuesta verbal y qué habría que observar para decidir entre ellas.

La Tabla 3.1 reúne los operantes que usa la ficha, con la variable de control de cada uno (Skinner, 1957) y un ejemplo construido para este capítulo, que no viene de ninguna fuente y sirve solo para fijar la diferencia.

TABLA 3.1

Operantes verbales: lo que controla la respuesta y un ejemplo construido en el discurso político

Operante	Lo que controla la respuesta (Skinner, 1957)	Ejemplo construido
Mando	Privación o estimulación aversiva del hablante: "la respuesta es reforzada por una consecuencia característica" y "no tiene una relación especificada con un estímulo previo" (pp. 35–36)	"Voten por mí."; "Basta de rodeos."; "Vamos a cambiar Brasil"
Tacto	Objeto, evento o propiedad de un objeto o evento: "en presencia del objeto o evento, una respuesta de esa forma es característicamente reforzada en una comunidad verbal dada" (pp. 81–82)	"El candidato empujó a su adversario durante el debate", dicho ante lo ocurrido
Ecoica	Estímulo verbal auditivo: "la respuesta genera un patrón sonoro similar al del estímulo" (p. 55). Hay correspondencia fonética entre el estímulo y las respuestas vocales	Repetir ("Bandido bueno es bandido muerto"), al responder, la expresión exacta usada por el adversario ("Bandido bueno es bandido muerto")
Textual	Estímulo verbal no auditivo (texto): "una respuesta vocal está bajo el control de un estímulo verbal no auditivo" (p. 66). Emisión de respuestas vocales con fonemas correspondientes a estímulos escritos	Leer en el pleno el pasaje del proyecto de ley
Transcripción	Estímulo verbal, con "una correspondencia punto por punto entre sistemas dimensionales diferentes" (p. 71). Copia: emitir una respuesta escrita con letras correspondientes al estímulo escrito. Dictado: emitir una respuesta escrita con letras correspondientes a los fonemas presentes en el estímulo vocal	Copia: teclear en el celular una etiqueta correspondiente a la generada por una IA en la computadora. Tomar dictado: tomar nota de las afirmaciones de un adversario.

Operante	Lo que controla la respuesta (Skinner, 1957)	Ejemplo construido
Intraverbal	Estímulo verbal sin correspondencia punto por punto con la respuesta: "algunas respuestas verbales no muestran correspondencia punto por punto con los estímulos verbales que las evocan" (pp. 71–72)	Responder "jamás será vencido" ante el inicio de la consigna "El pueblo unido"

Nota. Las definiciones son de Skinner (1957), en las páginas indicadas; la columna de ejemplos es del autor. Algunas categorías fueron complementadas por el autor, ya que la definición más precisa de Skinner se da a partir de fragmentos en párrafos. Los seis tipos son operantes definidos por las variables de control, no por formas: la misma frase puede pertenecer a más de un tipo en ocasiones diferentes.

3.3 ¿Tacto o intraverbal? El continuo y la forma de tacto

Solo en cuanto al tacto tiene sentido evaluar la veracidad o no de la "respuesta verbal" (dicha, escrita o gesticulada), ya que su variable de control es un estímulo no verbal, es decir, eventos del mundo. Skinner lo define como el operante "en el que una respuesta de forma dada es evocada (o al menos fortalecida) por un objeto o evento particular o por una propiedad de un objeto o evento" porque "en presencia del objeto o evento, una respuesta de esa forma es característicamente reforzada en una comunidad verbal dada" (pp. 81–82). La comunidad verbal lo establece porque los tactos aumentan el contacto de los oyentes con el ambiente. Cuando alguien afirma que un candidato a gobernador le dio un sillazo a otro durante un debate, está emitiendo un tacto siempre que el sillazo haya ocurrido de hecho. Paralelamente, un oyente que no vio el debate tendrá acceso a lo ocurrido (más específicamente, tendrá acceso al relato de lo ocurrido) mediante la emisión del tacto por el hablante. En el mando, a su vez, el oyente tiene acceso a lo que puede reforzar la respuesta verbal del hablante, ya que el mando es el único operante verbal cuyas respuestas especifican el reforzador y, en muchos casos, la acción del oyente.

CONCEPTO EN USO

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

Cualquier evento que suele preceder a muchos reforzadores diferentes puede funcionar como reforzador para la conducta en todas las condiciones de privación y de estimulación aversiva apropiadas; un ejemplo común es la aprobación, que puede ser poco más que un gesto de asentimiento o una sonrisa de quien suele proporcionar una variedad de reforzadores (Skinner, 1957). Como este reforzador no depende de un estado particular de privación, la forma de la respuesta pasa a depender del estímulo presente: así se establece el tacto, y por eso el reforzamiento generalizado es, para Skinner, la clave del discurso práctico y científico. Para Guerin (1994), las actitudes son enunciados mantenidos por

este reforzamiento social generalizado de una comunidad verbal. Como preceden a diversos otros reforzadores, los reforzadores generalizados no guardan una relación de correspondencia con topografías de respuesta específicas, es decir, refuerzan una infinita multiplicidad de respuestas. La aprobación, ilustrada por la expresión "¡Muy bien!", reforzará diferentes respuestas, como resolver una ecuación matemática o pronunciar correctamente una palabra en alemán, por ejemplo.

La pureza del tacto es, sin embargo, una cuestión de grado. Skinner llama "puro" u "objetivo" al tacto mantenido con reforzamiento totalmente generalizado, lo que permite el control exclusivo por el estímulo antecedente no verbal, algo que raramente sucede. En términos ideales, el discurso científico se ocupa de crear condiciones (método y rigor científicos) para que los tactos científicos se acerquen a ese ideal. Aun en el discurso científico ocurren distorsiones, ya que las comunidades científicas refuerzan ciertos relatos de resultados de investigación más que otros. Sin contar la ausencia de consenso epistemológico. En el ámbito político, a su vez, las distorsiones son mucho más probables, ya que nos referimos al control sobre la conducta de otras personas, en este caso, de los oyentes.

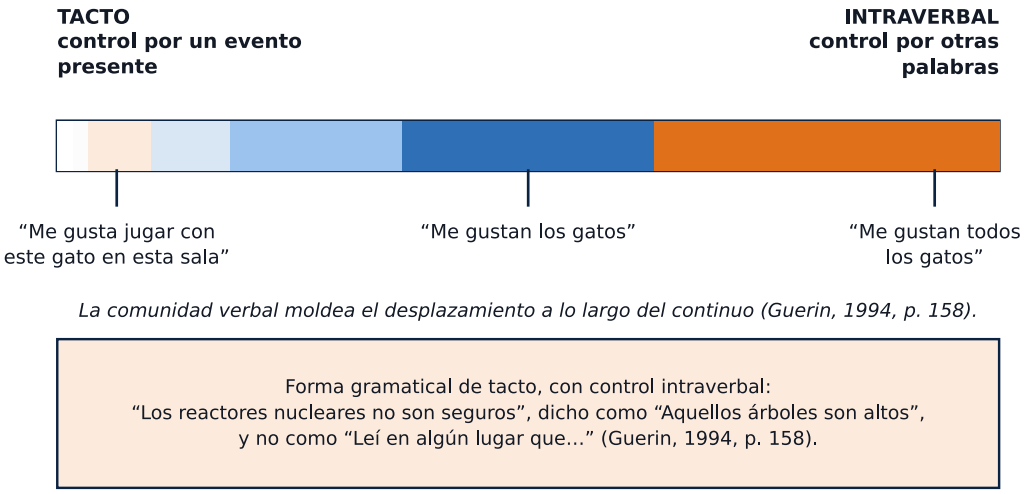
El intraverbal, en cambio, aunque sea reforzado por el mismo reforzador que el tacto, está controlado por estímulos verbales, que no guardan una relación de correspondencia punto por punto entre el estímulo y la respuesta. La inmensa mayoría de los contenidos enseñados en la educación formal son intraverbales. Tomando la historia, por ejemplo, aprendemos sobre el saqueo del rey Juan VI a las arcas brasileñas al regresar a Portugal leyendo u oyendo acerca de ese evento histórico, ya que no lo presenciamos. Lo que decimos o escribimos sobre ese evento se consideran respuestas intraverbales. Guerin (1994) lleva esta idea a la conducta de relatar creencias y actitudes. Para él, la distinción entre tactos e intraverbales es particularmente desafiante en varios ejemplos concretos, algo que el Análisis de la Conducta todavía no ha esclarecido satisfactoriamente. Las comunidades verbales moldean el desplazamiento a lo largo de un continuo, de "me gusta jugar con este gato en esta sala", con alguna plausibilidad de tacto, a "me gustan los gatos" y "me gustan todos los gatos", que son intraverbales. Muchas actitudes versan sobre eventos que la persona nunca experimentó, y, por lo tanto, lo que se sabe viene de los periódicos y de otras personas. La Figura 3.3 lo resume.

Lo que hace relevante el intraverbal para quien estudia la desinformación es lo que la comunidad refuerza. Para Guerin (1994), las comunidades verbales refuerzan la emisión de intraverbales como si fueran tactos (pp. 155, 158): "los reactores nucleares no son seguros", en la misma forma gramatical que "aquellos árboles son altos", es un enunciado más poderoso e influyente que "leí en

algún lugar que los reactores nucleares no son seguros" (p. 158). Guerin (1992) analiza el conocimiento socialmente construido como intraverbales, tactos de discriminación laxa y conducta verbal bajo fuertes controles sociales del grupo. Una misma frase, sin autoclíticos que añadan información sobre las variables de control (p. ej., "andan diciendo por ahí que..."), puede ser tanto un intraverbal como un tacto según la variable de control. Los tactos, en general, tienen más credibilidad entre los oyentes que los intraverbales. La oración "Vi a un determinado político casado entrando en un burdel" (es decir, el autoclítico "vi" especifica la respuesta verbal como un tacto) tiene más credibilidad que "Me dijeron que un determinado político casado entró en un burdel" (es decir, el autoclítico "me dijeron" especifica la respuesta como un intraverbal). Por otro lado, la oración "Un determinado político casado entró en un burdel" puede ser tanto un tacto como un intraverbal, porque no contiene elementos autoclíticos que informen al oyente su variable de control.

FIGURA 3.3

El continuo tacto–intraverbal y la forma de tacto



Nota. Esquema del autor, con los ejemplos de Guerin (1994). Guerin da "me gustan los gatos" y "me gustan todos los gatos" como intraverbales; la posición intermedia es ilustrativa. No representa datos.

Un ejemplo muestra lo que esto significa para el análisis. Mota et al. (2023) estudiaron diez pronunciamientos de Bolsonaro sobre la vacuna CoronaVac, entre julio de 2020 y enero de 2021. Eligieron, en buscadores, las diez declaraciones de mayor repercusión y de mención directa a los inmunizantes y registraron de cada una el número, la fecha, el contexto (medio y ocasión) y el contenido transcrito. Clasificaron los desórdenes informativos según la tipología de Wardle y Derakhshan, en la que desinformación es información falsa creada deliberadamente para perjudicar e información incorrecta es información fal-

sa sin intención de causar daño; y la mala información es la presentación de información verdadera, pero fuera de contexto, para causar daño. El tipo más frecuente fue la información incorrecta (47,1%), seguida de la desinformación (35,3%) y de la mala información (17,6%). En una de las declaraciones, el 29 de octubre de 2020, en la "live" semanal, Bolsonaro dijo que la vacuna que menos tiempo tardó en ponerse a disposición de la población llevó cuatro años. Las autoras leen la frase como información parcialmente verdadera (de hecho, la vacuna de desarrollo más rápido hasta entonces, contra las paperas, había llevado cuatro años) usada en un contexto falso, con la intención, según ellas, de poner en duda las vacunas contra la covid-19, cuyo desarrollo llevó menos tiempo. Lo que la ficha añade a la lectura de las autoras es la separación entre lo que el registro muestra y lo que ella infirió. El registro muestra una frase que, tomada sola, corresponde a un dato verificable. No muestra si la respuesta fue controlada por el hecho (y entonces la falla estaría en el contexto en que se usó) o por una secuencia de palabras aprendida como argumento (y entonces el hecho sería tangencial). Decir "contexto falso", como Mota et al., es describir el efecto de la frase sobre el oyente, y no la variable que la controló.

3.4 Tacto distorsionado: cuando la consecuencia del oyente desplaza el control por el evento

Skinner trata de la distorsión del tacto en el capítulo sobre las condiciones especiales que afectan el control de estímulos (1957). El punto de partida es que el reforzamiento generalizado es la clave del discurso práctico y científico: lleva la conducta del hablante al control del ambiente actual y permite al oyente reaccionar a lo que dice el hablante en lugar del contacto directo. Cuando la correspondencia con la situación estimulante se mantiene con nitidez, la respuesta se llama objetiva, válida, verdadera o correcta. Pero el control de estímulos nunca es perfecto, y puede ser distorsionado por consecuencias especiales que proporcionan un oyente particular, o los oyentes en general, caso en el que la respuesta se llama subjetiva, prejuiciosa o desiderativa. Cuando el reforzamiento generalizado es contingente a topografías de respuestas verbales específicas, correspondientes o no a los estímulos a los que se refieren, los tactos distorsionados se vuelven más frecuentes (Ribeiro, 1989). Si una persona dice que un candidato a gobernador es corrupto, aunque no tenga ninguna evidencia de ello, probablemente los opositores de ese candidato reforzaron esa respuesta verbal con aprobación y apoyo.

La descripción de Skinner es gradual, y el gradiente importa. En el caso menor, el hablante exagera el tamaño del pez que pescó o minimiza el peligro de un ataque; la respuesta todavía está controlada, en parte, por el evento. En un caso mayor, el control es inventado, y la respuesta se emite cuando el evento

no ocurrió: el niño que perdió una moneda, dijo que la perdió y fue atendido pasa a decirlo cuando no ha perdido ninguna, y el muchacho que gritó "¡Lobo!" es el estereotipo. El ejemplo más claro del mecanismo es el del trovador. Empieza relatando hazañas reales; ciertas partes reciben una aprobación especial porque interesan a los oyentes o los halagan; esas partes sobreviven en las narraciones siguientes, y bajo el mismo reforzamiento diferencial él pasa a estirar el relato, exagerando el tamaño de la batalla y el heroísmo de los participantes, y, por último, rompe por completo con el control por el estímulo, "describiendo" escenas que nunca vio. La distorsión es así producto de la selección por consecuencias: el oyente refuerza unas versiones, el hablante las repite y, paso a paso, la correspondencia deja de ser lo que decide.

CONCEPTO EN USO

Tacto distorsionado y subcomunidad verbal (Skinner, 1957; Guerin, 1992)

Tacto distorsionado es el tacto cuyo control por el evento fue desplazado por consecuencias especiales proporcionadas por el oyente. Skinner (1957, p. 149) describe grados: el hablante "estira los hechos"; el control es "inventado", y la respuesta se emite en ausencia de las circunstancias en que es característicamente reforzada; en el límite, la respuesta se emite en circunstancias que controlan una respuesta incompatible, y a eso se le llama mentira. Para Guerin (1992), el control del tacto puede estar en manos de una *subcomunidad verbal* menor: si un grupo pequeño refuerza ciertos relatos y castiga otros, las características del grupo pasan a controlar lo que se relata, aunque las consecuencias sean generalizadas, y Skinner llamó distorsionados a esos tactos.

Skinner describe además una mezcla que ocurre con frecuencia, el tacto impuro. Cuando un ama de casa dice "La cena está lista" principalmente porque los oyentes vendrán entonces a la mesa, la respuesta es funcionalmente muy próxima al mando "¡Vengan a cenar!"; la mezcla de relaciones de control características del tacto y del mando es lo que él llama tacto impuro. La Figura 3.4 muestra la lógica. El punto es que una respuesta puede ser función de las dos variables al mismo tiempo, y las fuentes de fuerza se suman. Esto importa en la interpretación de las declaraciones: aun cuando el evento controla parte de la respuesta, la consecuencia del oyente puede controlar otra parte, y la falsedad, medida por la correspondencia con el evento, no dice cuál es la proporción.

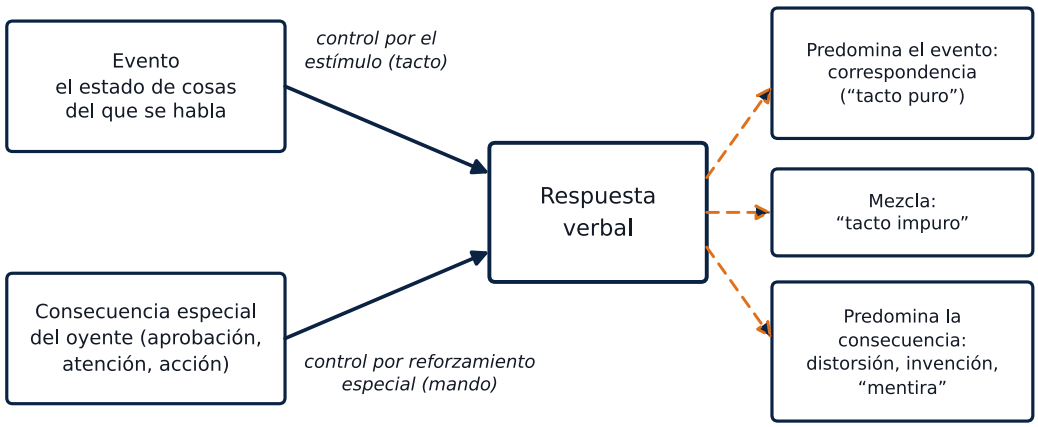
Skinner añade dos predicciones que interesan a la política. La primera es la del costo para el sistema. Cuando la distorsión lleva al oyente a reaccionar de modo ineficaz, el sistema hablante–oyente puede deteriorarse: el oyente puede suspender el reforzamiento o castigar las respuestas del hablante. El tacto distorsionado es "temporalmente más eficaz" porque explota una mayor ten-

dencia del oyente a responder; la comunidad deja de dar la moneda y puede castigar al niño por mentir. La segunda es la de la condición para la estabilidad: el sistema solo es estable cuando la correspondencia con los estímulos de control no tiene ninguna importancia práctica para el oyente, como en la literatura. Y está la observación sobre la aversión: el control de un tacto es especialmente propenso a la distorsión cuando la respuesta se emite para evitar consecuencias aversivas o escapar de ellas; el testigo sobornado es el ejemplo.

Estas predicciones fueron escritas para una díada (un hablante, un oyente) y para una comunidad en la que el oyente reacciona a la declaración con acciones que dependen de que corresponda a los hechos. La declaración política difiere en tres puntos, y es en esos puntos donde el análisis avanza por hipótesis. Primero, las audiencias son varias, y quien refuerza la distorsión puede no ser quien la castiga (sección 3.1); el deterioro previsto para la díada puede no ocurrir mientras la conducta del hablante sea reforzada por una audiencia y castigada solo por otra, sin acceso a ella. Segundo, para una parte de la audiencia, la correspondencia con los hechos puede no tener, en la terminología de Skinner, ninguna importancia práctica, y la función de la frase pasa a ser otra: los efectos que produce en los oyentes. Guerin (1994) señala que relatar actitudes puede servir de evento social ritual, en el que el reforzamiento está en el propio intercambio. Tercero, cuando el hablante pertenece a una subcomunidad que refuerza ciertas respuestas y castiga otras, lo que se tiene es lo que Guerin (1992) describe: lo que se relata pasa a depender del grupo. Los Capítulos 6 y 9 tratan de las dos últimas.

FIGURA 3.4

Control múltiple: el evento y la consecuencia especial del oyente actúan sobre la misma respuesta



Una única respuesta puede ser función de las dos variables al mismo tiempo; las fuentes de fuerza se suman (Skinner, 1957, pp. 227–228).

Nota. Esquema del autor, a partir de Skinner (1957). Los tres desenlaces son tipos, no proporciones; ninguna fuente mide cuánto de la fuerza de una respuesta política viene de cada variable.

DATO	Skinner (1957) describe, sin medir, tres propiedades del tacto distorsionado: es función de consecuencias especiales del oyente; es temporalmente más eficaz que el tacto; y lleva al sistema hablante–oyente a deteriorarse, salvo cuando la correspondencia no le importa al oyente. Guerin (1992) describe el desplazamiento del control hacia grupos pequeños.
INTERPRETACIÓN	La declaración política puede leerse con la misma lógica, con las salvedades de las audiencias múltiples y de las funciones no informativas. Las declaraciones clasificadas como falsas por una agencia son candidatas al análisis de distorsión, no casos demostrados de ella.
HIPÓTESIS	En un campo en el que la correspondencia con los hechos tiene poca importancia práctica para la audiencia que refuerza, el tacto distorsionado no se deteriora y tiende a estabilizarse. Ninguna fuente verificada mide esto para declaraciones de políticos.

3.5 Repetición: la fuerza de la afirmación repetida

La repetición aparece en dos posiciones y conviene no confundirlas. Está la repetición del hablante, que dice de nuevo la misma frase en ocasiones diferentes (lo que la configura como respuesta intraverbal), y la repetición que sufre el oyente, cuando se encuentra de nuevo con la misma frase. La primera es variable dependiente del análisis operante: la frecuencia de emisión es medida de la fuerza de la respuesta. La segunda es variable independiente en buena parte de la psicología de la desinformación.

Del lado del hablante, Skinner ofrece tres puntos de apoyo. El primero es la observación de que una anécdota o un chiste particularmente exitoso tiende a contarse de nuevo (1957, p. 148): las medidas especiales de reforzamiento "le dicen al hablante de qué vale la pena hablar" (p. 148). El segundo es la advertencia de que la repetición puede no ser más que evidencia de una fuerza excesiva de la respuesta, y no de una variable de control específica. El tercero es que, cuando la conducta verbal es ineficaz para alterar las circunstancias responsables de su fuerza, una respuesta puede ceder ante otra, sobre todo cuando la repetición pura es castigada, y lo que aparece es una "repetición con variación" (p. 228). La predicción es comprobable: si la repetición pura es castigada, la forma debe variar. Esto ocurre mucho en la redacción, en la que los profesores castigan las frases repetidas a lo largo de un texto.

Lo que las fuentes verificadas documentan para Brasil es el número de veces que ciertas afirmaciones se repitieron. Aos Fatos, agencia brasileña de verificación de datos, registró que la afirmación de que los gobiernos del PT (Partido dos Trabalhadores) habrían acabado con el hambre en Brasil se dijo en el discurso de investidura de Lula, en el balcón del Palácio do Planalto (sede de la Presidencia), y se repitió en al menos otras siete ocasiones en 2023, y que era recurrente en su discurso desde la campaña (Ribeiro, 2023); en 2024, la afirmación fue la más repetida del año, en al menos 24 ocasiones, y Aos Fatos la clasificó como falsa por comparar dos indicadores distintos (Menezes, 2024). La afirmación de que sería el único presidente sin título universitario se repitió en al menos 17 ocasiones en 2024 y fue clasificada como falsa (Menezes, 2024). En el otro campo, la fuente equivalente es de naturaleza diferente. Mota et al. (2023) transcriben dos declaraciones de Bolsonaro, del 17 de diciembre de 2020 y del 15 de enero de 2021, que repiten, con palabras muy próximas, dos cosas: que él ya había contraído el virus y tenía anticuerpos ("¿para qué vacunarme de nuevo?", "¿para qué vacunarme?") y que los laboratorios, en los contratos, no se responsabilizan por los efectos colaterales; y, según las autoras, que citan un reportaje del diario O Globo de enero de 2021, él hizo al menos diez críticas públicas a la CoronaVac entre julio de 2020 y enero de 2021.

Cuatro observaciones separan lo que esto muestra de lo que no muestra. La primera: las dos fuentes miden cosas diferentes (Aos Fatos cuenta las ocasiones en que una afirmación se repitió en un año, a partir de discursos, entrevistas y transmisiones en vivo de 2023 y, en 2024, de todos los discursos, pronunciamientos y entrevistas del presidente; las autoras transcriben diez declaraciones elegidas por su repercusión), y por eso los recuentos no son comparables. La segunda: la repetición de la afirmación sobre el hambre continuó en 2024 después de que el relevamiento de 2023 ya la hubiera clasificado como falsa (Ribeiro, 2023; Menezes, 2024); es un dato que no dice que Lula conociera la clasificación, ni que la audiencia la conociera. La tercera: lo que se registra es la emisión, y no lo que mantiene la emisión. La cuarta: las repeticiones registradas aquí son, en la superficie, repetición pura, y el capítulo no sabe si en otras ocasiones la forma varió.

Del lado del oyente, la fuente verificada es un conjunto de experimentos. Effron y Raj (2020) reunieron cuatro experimentos y un piloto, dos de ellos preregistrados, con 2.587 participantes en total, para preguntar si encontrarse de nuevo con un titular falso cambia el juicio moral sobre divulgarlo. En el Experimento 1, 138 estadounidenses de un panel en línea vieron seis de 12 titulares falsos cuatro veces, en una fase de familiarización en la que evaluaban, cada vez, un aspecto sin relación con la falsedad; después juzgaron los 12, seis

ya vistos y seis nuevos, informados de que verificadores no partidistas habían confirmado que esos hechos no ocurrieron. Haber visto el titular una o cuatro veces redujo cuán poco ético les parecía a los participantes publicarlo y compartirlo cuando lo veían de nuevo, aun rotulado como falso y sin que lo creyeran, e incluso después de descontar estadísticamente cuán simpático y popular parecía; y percibir el titular como menos antiético predijo una mayor inclinación a expresarle aprobación en línea. En un ambiente experimental, las personas compartieron más los titulares repetidos que los nuevos (Effron & Raj, 2020). Los autores especulan que la repetición puede reducir la condena moral al hacer que la información parezca intuitivamente verdadera, y discuten otros mecanismos. En términos operantes, el resultado es compatible con una lectura: la repetición cambia lo que la audiencia hace con la frase (condena menos, comparte más), es decir, cambia la consecuencia que el hablante recibe por la repetición. Es una lectura del capítulo, no una conclusión de los autores. Y el diseño (titulares, participantes de experimentos, ambiente simulado) está lejos de un discurso de campaña.

DATO	Una afirmación fue repetida en al menos 24 ocasiones en 2024 por Lula (Menezes, 2024), después de haber sido clasificada como falsa en 2023 (Ribeiro, 2023); dos cosas dichas por Bolsonaro el 17/12/2020 reaparecen el 15/01/2021 (Mota et al., 2023). En experimentos, la repetición redujo la condena moral y aumentó las veces que se compartía (Effron & Raj, 2020).
INTERPRETACIÓN	La repetición es una medida de la fuerza de la respuesta, y la fuerza es efecto de la historia de reforzamiento; la literatura experimental sugiere que la audiencia tolera más lo que ya oyó. Los registros no identifican qué consecuencias mantuvieron las repeticiones.
HIPÓTESIS	La repetición por parte de los políticos se mantiene, en parte, por la reacción de la audiencia a ella, y debe variar en la forma cuando la repetición pura pasa a ser castigada. Para comprobarlo, hace falta el registro de las reacciones de la audiencia en cada ocasión y de la forma de la afirmación antes y después de la verificación.

3.6 Autoclíticos de fuente y de certeza: lo que se añade y lo que se quita

En el Capítulo 2 el autoclítico apareció como una calificación del enunciado que modifica su efecto sobre el oyente (Guerin, 1994). Skinner (1957, cap. 12) lo analiza por entero, y tres familias interesan aquí.

La primera familia es la de los autoclíticos de fuente. Skinner señala que preguntar "¿Lo vio usted, o alguien se lo contó?" es preguntar si la respuesta fue tacto, ecoica o intraverbal a la conducta verbal de otra persona, y que, en muchas comunidades, es de buena educación empezar a hablar con un autoclítico que indique de dónde viene lo que se dice; un autoclítico puede indicar además que lo que se dice ya es aceptado por los oyentes o por las personas en general. Son el contrapeso de lo que describió la sección 3.3: la comunidad refuerza intraverbales con forma de tacto, y el autoclítico de fuente es justamente lo que diría que la respuesta no es tacto. La Figura 3.5 propone la hipótesis de que, a lo largo de una cadena de retransmisión, los autoclíticos de fuente y de debilidad se pierden y lo que queda es la forma de tacto; Skinner describe que una respuesta atípica en una conversación tiende a ser retomada y pasada de hablante en hablante, pero ninguna fuente verificada sigue una cadena de ese tipo.

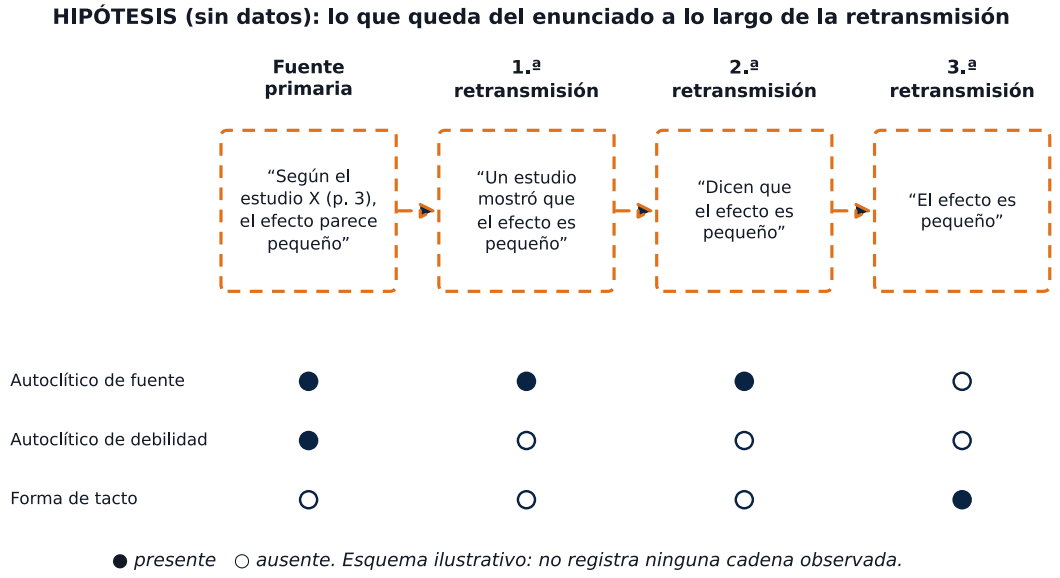
CONCEPTO EN USO

Autoclíticos de fuente y de certeza (Skinner, 1957)

El autoclítico es conducta verbal basada en otra conducta verbal del mismo hablante y que modifica la reacción del oyente a ella. Los *descriptivos* informan el tipo de operante que acompañan ("veo que…", para lo que se leyó; "recuerdo") y la fuerza de la respuesta: "supongo" y "creo" indican estimulación insuficiente o una respuesta mal condicionada (p. 315); "sé", "le aseguro" y "juro" sugieren la fuerza de la respuesta, y podrían omitirse sin cambiar la naturaleza del efecto, pero lo hacen más preciso. Están los que indican la condición emocional del hablante ("me complace decir"; p. 316) y los que señalan la fuente ("dicen que…"). Los *calificadores* incluyen el "no", que tiene fuerza de mando y se traduce como "no actúe sobre esta respuesta como un tacto no extendido" (p. 324), y el autoclítico asertivo, que ordena al oyente que acepte un estado de cosas. Los *cuantificadores* ("todos", "algunos", "ninguno") modifican la reacción del oyente a la oración entera.

FIGURA 3.5

Hipótesis: pérdida de los autoclíticos de fuente y de debilidad a lo largo de una cadena de retransmisión



Nota. Hipótesis aquí propuesta. Las frases son construidas; el esquema no registra ninguna cadena observada y sirve para decir lo que habría que observar: el enunciado en cada eslabón y los autoclíticos de cada uno.

La segunda familia es la de los autoclíticos de debilidad y de certeza, y un ejemplo está en una de las declaraciones registradas por Mota et al. (2023). El 24 de diciembre de 2020, Bolsonaro dijo, en su transmisión semanal: "La eficacia de aquella vacuna en São Paulo parece que está por los suelos, ¿no? No voy a divulgar el porcentaje aquí, porque si me equivoco en un 0,001% los medios me van a caer encima, pero parece que el porcentaje está por los suelos" (p. 320). Las autoras clasifican como falsa la información de que el porcentaje estaría "por los suelos" (p. 323). La frase trae dos veces "parece que", que es un autoclítico de debilidad en el sentido de Skinner (indica una respuesta basada en estimulación insuficiente), y una referencia explícita a la audiencia que puede castigarla: los medios. La forma, por lo tanto, informa al oyente el grado de fuerza de la respuesta, y la referencia a los medios describe la contingencia que, según el propio hablante, lo controlaría si diera la cifra. La declaración no tiene la forma de tacto que describió la sección 3.3: se presenta como una conjetura. En la hipótesis de la Figura 3.5, es un enunciado así, con autoclítico de debilidad, el que perdería el autoclítico al ser retransmitido; el registro de Mota et al. no muestra si eso ocurrió.

La tercera familia es la más pertinente para los datos de Lupa. Los términos absolutos (“siempre”, “nunca”, “todos”, “ninguno”, “único”, “primero”, “mayor”, “menor”) y las negaciones fuertes (“no existe”, “no hay”) son, en la clasificación de Skinner, cuantificadores y calificadores: “todos” equivale a “siempre es posible decir” y “ninguno” a “nunca es posible decir” (p. 329), y el “no” tiene fuerza de mando. Definen la extensión con que se le pide al oyente que acepte la respuesta, y Skinner señala, a propósito del autoclítico asertivo (y no de los cuantificadores), que cualquier condición colateral que debilitaría la reacción del oyente, como la negación por otra persona o unas circunstancias dudosas, lleva al hablante a intensificarlo.

Lupa midió la frecuencia de estos patrones en las frases que verificó en períodos electorales, y el resultado merece una lectura cuidadosa. Según el informe, las afirmaciones con términos absolutos fueron clasificadas como “Falso” en el 40,6% de las verificaciones en período electoral, y las frases con negaciones del tipo “no existe” o “no hay”, en el 45,9% (Agência Lupa, 2026); las que mencionan el proceso electoral, en el 33,2%; las que atacan a adversarios, en el 27,9%; y las que traen números, porcentajes o valores, en el 18,3%, aunque estas reciben con más frecuencia “Exagerado”, “Subestimado” o “Verdadero, pero / falta contexto” (Verdadeiro, mas / falta contexto) (p. 50). De las 936 verificaciones clasificadas como “Exagerado” o “Verdadero, pero / falta contexto”, 737 (78,7%) contenían algún marcador de escala, comparación, cantidad, desempeño, autoría o legado. El texto no informa los numeradores y denominadores de los porcentajes de 40,6%, 45,9%, 33,2% y 27,9%, ni si la base de cada porcentaje es el conjunto de las verificaciones de cada patrón.

Hay al menos tres lecturas que no se excluyen. La lectura lógica: una frase universal o negativa es refutada por un único contraejemplo, y una frase con salvedad no lo es; la tasa de “Falso” sería, en parte, una propiedad de la forma lógica, y no del hablante. La lectura de selección: las agencias verifican lo que es verificable y lo que circula; Lupa reconoce que el volumen de verificaciones de cada político depende del tiempo de exposición, de las entrevistas y debates y, “sobre todo”, del estilo de comunicación, y que los datos reflejan solo lo que Lupa publicó en su sección de Elecciones, condicionado a decisiones editoriales y de relevancia. Y la lectura operante, que es la que aquí se propone: el autoclítico absoluto se emite donde la reacción del oyente necesita ser fortalecida, es decir, en ocasiones de disputa, de cuestionamiento o de duda. La tercera solo se comprueba con el registro de la ocasión, que el informe no trae.

3.7 La declaración como regla para la audiencia

Una declaración política puede tener otra función además de describir un estado de cosas: describir una contingencia, es decir, decirle al oyente lo que sucederá si hace esto o aquello. “Quien se ponga esa vacuna se va a convertir en caimán” describe un evento (el efecto de la vacuna), pero, sobre todo, una relación entre la conducta del oyente y una consecuencia. En la literatura analítico-conductual, ese operante y la conducta que evoca se tratan como reglas, y la conducta gobernada por reglas es la conducta del oyente en relación con ellas.

CONCEPTO EN USO

Regla y conducta gobernada por reglas (Baum, 1995; Cerutti, 1989)

Para Baum (1995), una regla es un estímulo discriminativo verbal producido por el hablante bajo el control de una contingencia, es decir, de una relación entre la conducta y sus consecuencias. Cerutti (1989) define la conducta gobernada por reglas como la combinación, en instrucciones complejas, de discriminaciones elementales establecidas antes. Se afirma que la conducta está gobernada por reglas cuando se demuestra que su probabilidad de ocurrencia es afectada por el contacto con una regla. La propiedad que interesa aquí es la *insensibilidad a las contingencias*: los desempeños instruidos son menos sensibles a los cambios en las consecuencias programadas (Hayes et al., 1986).

La literatura experimental disponible aquí, toda de laboratorio, es consistente con la idea de que la regla, en sí misma, puede controlar la conducta del oyente sin que este entre en contacto con la contingencia. En Galizio (1979), seis estudiantes de 18 a 25 años, reclutados mediante un anuncio de empleo de medio tiempo en el que el pago dependía del desempeño, giraban una palanca para evitar pérdidas de dinero, señaladas por una luz roja y un tono; cuatro luces ámbar, con rótulos, indicaban el componente del programa múltiple de evitación en vigor. Las instrucciones exactas controlaron el desempeño en cada componente, y el seguimiento de instrucciones se eliminó cuando instrucciones inexactas llevaron a los participantes a entrar en contacto con una contingencia de pérdida monetaria; mientras la pérdida programada estaba retirada, los participantes siguieron la instrucción inexacta, sin tendencia a la baja en las tasas. Hayes et al. (1986) presentaron a los participantes una tarea en la que apretar botones movía una luz por un laberinto, y mover la luz daba puntos que valían oportunidades de premios en dinero; el movimiento seguía un programa múltiple de razón fija 18 y de reforzamiento diferencial de tasas bajas de 6 segundos, con componentes que alternaban cada 2 minutos. En el segundo experimento, luces emparejadas con las instrucciones “Go Fast” y

"Go Slow" ("ve rápido", "ve despacio") se presentaban y después se retiraban. Los autores señalan que los desempeños instruidos son menos sensibles a los cambios en las consecuencias y proponen que parte de esa insensibilidad se debe a contingencias adicionales introducidas por el control instruccional, como las consecuencias sociales de seguir la regla.

Para el análisis de declaraciones, la consecuencia es doble. La primera es que la falsedad de una regla puede estar en la descripción de la contingencia, y no en la descripción del evento: "si te conviertes en caimán, es problema tuyo", registrada por Mota et al. (2023) en la declaración del 17 de diciembre de 2020, tiene una parte descriptiva (el efecto colateral) y una parte de regla (la consecuencia para quien se vacuna). Las autoras la leen como un probable recurso satírico o de hipérbole, que debe observarse junto al impacto potencial de la declaración de un jefe de Estado. Según la ficha, lo que se registra es la forma (regla, con consecuencia descrita) y lo que se plantea como hipótesis es la función: si el oyente que la sigue entra en contacto con la contingencia real, el experimento de Galizio sugiere que el seguimiento se extingue; si no entra en contacto, o si seguir la regla tiene consecuencias sociales, la insensibilidad descrita por Hayes y colegas hace resistente el seguimiento de la regla. La segunda es la dependencia de la audiencia: la regla falsa resiste en la medida en que el oyente no entre en contacto con la contingencia real y en que otras contingencias, las sociales, mantengan el seguimiento de la regla. Ninguna fuente verificada mide esto con declaraciones políticas.

3.8 Casos: las mismas cuatro frases, con el mismo método

La Tabla 3.2 aplica la ficha a las cuatro frases de la apertura. Dos son de Lula y dos de Bolsonaro, todas de los ejemplos que Lupa reproduce en el informe, todas clasificadas por ella como "Falso", con el mismo método y en la misma publicación. La ficha repite, para cada una, lo que la fuente documenta y lo que no documenta. La columna del operante es interpretación, en los términos de la sección 3.2.

TABLA 3.2

Cuatro declaraciones clasificadas como "Falso" por Lupa: dato, interpretación y lo que la fuente no documenta

Caso	Declaración, fecha, contexto y clasificación (dato)	Interpretación (operante)	No documentado
1. Lula	"La menor deforestación de la Amazonía se hizo en mi gobierno." Debate de Band, Folha, TV Cultura y UOL, 28/08/2022. Lupa: "Falso"; la menor tasa registrada (4,6 mil km², 2012) fue en el gobierno de Dilma Rousseff (PT)	Tacto sobre el propio gobierno, con superlativo; distorsión en la dirección del mérito, o intraverbal de una narrativa sobre "los gobiernos del PT"	Pregunta o argumento al que respondía; reacción del público; si el dato ya había sido verificado
2. Lula	"Estoy orgulloso de haber nombrado al primer negro para la Corte Suprema." Mismo debate. Lupa: "Falso"; dos ministros negros habían integrado el STF antes del nombramiento de Joaquim Barbosa, en 2003	Autoclítico emocional ("estoy orgulloso") más un tacto con ordinal; hecho histórico, en general adquirido como intraverbal, emitido con forma de tacto	Ídem; si había una pregunta sobre el tema; reacción
3. Bolsonaro	"Ningún país del mundo compró vacunas en 2020." Debate de TV Globo, 29/09/2022. Lupa: "Falso"; al menos 50 países empezaron a vacunar en 2020 (Poder360); en Brasil, la vacunación empezó el 17/01/2021	Tacto con cuantificador universal negativo; posible función de tacto impuro, según lo que el oyente haría con la frase (hipótesis; la ocasión no está documentada)	La pregunta o el argumento anterior; si la frase era recurrente; reacción
4. Bolsonaro	"Yo nunca fui homofóbico." Entrevista al programa Central das Eleições (GloboNews), 03/08/2018. Lupa: "Falso"; condena por daño moral colectivo confirmada por el TJRJ en nov./2017, por declaraciones hechas en 2011 en el programa CQC	Tacto de la propia conducta verbal, con "nunca"; posible distorsión bajo contingencia aversiva (hipótesis)	La pregunta que la ocasionó; si la frase se repitió; reacción; respuesta a las correcciones

Nota. Agência Lupa (2026). Las frases, la clasificación y las justificaciones están en imágenes de las verificaciones reproducidas en el informe (recortes en verificacao/evidencias/); la columna de interpretación es del autor. "Falso" es la etiqueta que Lupa aplica a la información comprobadamente incorrecta.

Los casos de Lula tienen en común la primera persona y el mérito: “en mi gobierno”, “estoy orgulloso de haber nombrado”. El informe los presenta como ejemplos de un estilo, y dice que una parte considerable de las declaraciones “Falsas” de Lula en campaña está relacionada con la autoría o con el carácter pionero de políticas públicas, programas, leyes e hitos institucionales (Agência Lupa, 2026). Para la ficha, importa que, en el caso 1, la justificación de Lupa señala el dato verdadero (la menor tasa de la serie) en un gobierno del mismo partido, de modo que el enunciado difiere del dato en una sola variable: a quién se atribuye. Hay al menos dos lecturas. En una, la consecuencia del oyente (aprobación por los méritos) desplazó el control: es el tacto distorsionado en la dirección del mérito. En la otra, el hablante emitió un intraverbal, aprendido dentro de una narrativa sobre “los gobiernos del PT”, y la primera persona es solo la forma que esta tomó en ese momento. Se distinguen por dos observaciones que el informe no trae: lo que el mismo hablante dice del mismo dato ante audiencias que no refuerzan el mérito, y si la forma cambió después de la verificación. En el caso 2, la frase tiene dos partes: la justificación de Lupa reconoce el nombramiento de Joaquim Barbosa por Lula en 2003 y señala dos antecesores negros en el STF, de modo que la falla está en “primero”. “Estoy orgulloso de” es, en el vocabulario de Skinner, un autoclítico que indica la condición emocional del hablante y enfatiza la relación personal con él; el ordinal es el punto de falla.

Los casos de Bolsonaro, en el informe, ilustran otro estilo: Lupa dice que suele recibir “Falso” cuando usa números, porcentajes y estadísticas, y cuando hace afirmaciones categóricas y acusaciones contra adversarios, con “nunca”, “jamás”, “ninguno” y “siempre” apareciendo con frecuencia. Ambos casos tienen un cuantificador absoluto. En el caso 3, el contraejemplo es público y verificable (el Reino Unido empezó a vacunar el 8 de diciembre de 2020, según Lupa), que es exactamente lo que hace arriesgada la forma universal: una única excepción la refuta. En el caso 4, la frase es un tacto de la propia conducta verbal, y Skinner (1957) escribe que las contingencias necesarias para la conducta autodescriptiva las dispone la comunidad cuando tiene razones para preguntar “¿Qué dijo usted?” y “¿Por qué dijo eso?” (p. 314). Guerin (1994) añade que una pregunta puede funcionar como mando de la respuesta. La fuente no trae la pregunta, y es la pregunta la que decide si la respuesta se emite bajo la contingencia de una acusación, que es la condición en la que, según Skinner, el tacto es especialmente propenso a la distorsión. La ficha deja el campo en blanco, y la lectura como distorsión bajo contingencia aversiva queda como hipótesis.

Hay otras declaraciones de Bolsonaro documentadas con fecha, contexto y transcripción gracias a las diez que analizaron Mota et al. (2023), y entraron en las secciones 3.3, 3.5, 3.6 y 3.7. Entre las fuentes verificadas no hay un equivalente para Lula: ninguna transcribe declaraciones suyas con fecha y contexto y las clasifica con el mismo esquema. Esto es un límite del acervo de fuentes leídas y no una propiedad de los políticos, y debe leerse así: una conclusión comparativa a partir de esa asimetría sería un artefacto de la documentación. Mota et al. clasifican las declaraciones según la tipología de Wardle y Derakhshan (sección 3.3), en la que la desinformación se distingue de la información incorrecta por la intención de causar daño. La clasificación, en ese esquema, depende de una inferencia sobre la intención que el registro de una declaración no contiene; la ficha la sustituye por una pregunta sobre las variables de control.

La Tabla 3.3 reúne lo que el informe de Lupa publica sobre los patrones de lenguaje en las frases verificadas de cada uno, con la misma taxonomía y el mismo formato.

TABLA 3.3

Patrones de lenguaje en las frases verificadas por Lupa en campañas: Lula y Bolsonaro

Patrón en la frase verificada	Lula: frases (falsas; %)	Bolsonaro: frases (falsas; %)
Negaciones fuertes	4 (3; 75,0%)	25 (15; 60,0%)
Términos absolutos	7 (4; 57,1%)	39 (21; 53,8%)
Ataques a adversarios o partidos, delito o corrupción	12 (7; 58,3%)	52 (21; 40,4%)
Proceso electoral, voto y urna	3 (1; 33,3%)	13 (6; 46,2%)
Números, porcentajes o valores	57 (7; 12,3%)	94 (37; 39,4%)
Total de frases verificadas	83 (22; 26,5%)	236 (102; 43,2%)

Nota. Agência Lupa (2026, p. 25), "Taxonomia Lupa", valores leídos en la imagen de la página (recorte en verificacao/evidencias/lupa2026-p25-estilo-lula-bolsonaro.png). Las categorías no son excluyentes: una misma frase puede contener un número, un término absoluto y un ataque al adversario. Solo entran políticos con más de 30 frases verificadas; las frases son de campañas electorales, y Lupa no distingue los ciclos de Lula.

El capítulo registra lo que la tabla permite y lo que no permite. Permite decir que, en las frases que Lupa verificó, las que tenían negación fuerte, término absoluto o ataque al adversario recibieron "Falso" en mayor proporción que las que traían números, en los dos políticos, y que, entre las frases con números, la proporción de "Falso" fue del 12,3% (7 de 57) para Lula y del 39,4% (37 de 94) para Bolsonaro. No permite decir que uno de ellos tienda más que el otro a afirmar lo que es falso: las bases son selecciones de la agencia, con tamaños muy diferentes, las categorías se superponen, las celdas de Lula tienen 3, 4, 7 y 12 frases, y la propia Lupa dice que no hace un ranking de quién

"mintió más", porque el número de verificaciones depende del tiempo de exposición, del cargo y del estilo de comunicación. En estos términos, el dato es la clasificación de la agencia; la interpretación es que la forma de la frase se correlaciona con la clasificación; la hipótesis es la de la sección 3.6.

3.9 Interpretaciones rivales y cómo decidir entre ellas

Para una declaración clasificada como falsa, la ficha trabaja con seis interpretaciones rivales, que la Tabla 3.4 reúne. No se incluye entre ellas la "intención de engañar", porque la intención no es una variable de control observable; lo que hay son las consecuencias para la conducta del hablante, su historia de reforzamiento y las propiedades de la ocasión. La definición de mentira de Skinner es funcional: una respuesta emitida en circunstancias que normalmente controlan una respuesta incompatible (1957).

TABLA 3.4

Interpretaciones rivales para una declaración clasificada como falsa y lo que la ficha necesitaría registrar para separarlas

Interpretación	Variable de control en juego	Lo que la ficha necesitaría registrar
Control débil por el evento (error, lapsus)	Estimulación insuficiente o respuesta mal condicionada; sin consecuencia especial	Si la misma persona describe el evento correctamente en otras ocasiones, sin consecuencia especial; si hay autoclítico de debilidad
Intraverbal adquirido de otros	Estímulo verbal anterior (texto, discurso, mensaje)	La fuente de la cadena de palabras; si el enunciado reproduce una formulación anterior; ausencia de autoclítico de fuente
Tacto distorsionado	Consecuencia especial del oyente	Si la forma cambia con la audiencia; si la correspondencia es mayor cuando la audiencia no refuerza
Mando disfrazado o tacto impuro	Privación o aversión del hablante; lo que el oyente hará después	Lo que el hablante le pide al oyente, de modo explícito o implícito, en la misma ocasión (Capítulo 5)
Función ritual o de identidad	Reforzamiento social generalizado: el intercambio es el reforzamiento	Si el contenido varía sin cambiar el efecto; si la audiencia reacciona a la adhesión, y no al contenido (Guerin, 1994)
Efecto de la selección por la agencia	La declaración entró en la base por ser verificable y circular	La base y el método de selección de la agencia (Capítulo 2)

Nota. Recuadro del autor, a partir de Skinner (1957) y Guerin (1994). Las seis interpretaciones no se excluyen: Skinner (1957) señala que una respuesta suele ser función de más de una variable.

Hay cuatro tipos de observación que la ficha pediría y que los registros usados en este capítulo no traen. El primero es la ocasión: la pregunta, el argumento o el contexto inmediato al que responde la declaración. El segundo es la audiencia: quién estaba presente y cómo reaccionó, que es la consecuencia de la que depende el tacto distorsionado. El tercero es la comparación del mismo hablante y del mismo evento ante audiencias diferentes, o antes y después de la verificación; es el único modo de separar la distorsión, por consecuencia especial, del control débil por el evento. El cuarto es lo que sucede después de la corrección: si la forma permanece (repetición pura), varía (repetición con variación) o desaparece. Los Capítulos 6 y 9 tratan de lo que dicen las fuentes sobre la reacción de la audiencia y sobre lo que sucede después de la corrección; los Capítulos 5 y 7 toman de aquí la distinción entre tacto, mando e intraverbal para analizar el discurso de odio y el voto.

DATO	Cuatro declaraciones, dos de cada político, fueron clasificadas como “Falso” por Lupa; una serie de declaraciones de Bolsonaro sobre la vacuna fue transcrita y clasificada por Mota et al. (2023); dos afirmaciones de Lula fueron contadas como repetidas por Aos Fatos.
INTERPRETACIÓN	Las declaraciones son compatibles con más de un operante, y la forma gramatical de tacto y los cuantificadores absolutos aparecen en las cuatro. La clasificación de falsedad informa la correspondencia con un dato, y no la variable de control.
HIPÓTESIS	Parte de las declaraciones es función de consecuencias especiales de la audiencia (tacto distorsionado), parte de intraverbales aprendidos de otros y parte de funciones no informativas (mando, ritual). Para separarlas es preciso registrar la ocasión, la audiencia, la comparación y lo que ocurre después de la corrección.

Síntesis

Una declaración política es conducta verbal, y lo que la identifica como tacto, mando o intraverbal es lo que la controla, y no la forma. El tacto corresponde a un evento, y el intraverbal a otras palabras; la comunidad refuerza que el segundo se presente con la forma del primero. El tacto distorsionado es el tacto cuyo control por el evento fue desplazado por la consecuencia especial del oyente, en grados que van de la exageración a la invención. La repetición es una medida de fuerza, y la literatura experimental sugiere que la audiencia condena menos lo que ya oyó antes. Los autoclíticos de fuente y de certeza le dicen al oyente de dónde viene lo que se dice y con cuánta fuerza; los absolu-

tos y las negaciones le piden que acepte un estado de cosas, y Lupa muestra que las frases así reciben “Falso” con más frecuencia, por razones lógicas, de selección y, tal vez, operantes.

La ficha separa dato, interpretación e hipótesis. Aplicada a cuatro declaraciones de dos políticos, muestra lo que cada fuente documenta y lo que no documenta, y deja abierta la pregunta que importa: qué mantiene la declaración. Los Capítulos 5 y 7 toman de este capítulo la distinción entre tacto, mando e intraverbal; el Capítulo 6, el tacto distorsionado y la subcomunidad verbal; el Capítulo 9, el reforzamiento social generalizado y la audiencia; el Capítulo 12 (secciones 12.1 y 12.4), la idea de que corregir el enunciado es actuar sobre una respuesta cuyas variables de control pueden no ser las que el corrector supone.

Referencias

Agência Lupa. (2026, 27 de julio). *Anatomia da desinformação eleitoral* [Informe]. 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 diciembre). *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 diciembre). *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.

instituto WALDEN4 | ACADEMIA WALDEN4

WALDEN4 RESPONDE

La duda que llega al consultorio, respondida

Preguntas de quienes buscan terapia, de familias que recibieron un diagnóstico y de estudiantes, respondidas por quienes atienden en el día a día, con lo que dice la investigación y lo que todavía no sabe.

60+ respuestas publicadas

Acceda, gratis y sin registro

walden4.com.br/responde

CAPÍTULO

4

Actitudes políticas y prejuicio como clases de equivalencia

Muchas actitudes versan sobre cosas que la persona nunca ha experimentado. Guerin (1994) da el ejemplo de la energía nuclear: lo que la persona sabe sobre ella viene de los periódicos y de otras personas, no del contacto con el evento. En política, lo mismo parece valer para buena parte de lo que se evalúa, desde partidos e instituciones hasta grupos y políticos que el elector solo conoce por su imagen y su nombre. La pregunta que organiza este capítulo es cómo un nombre, una imagen o un símbolo llega a tener función evaluativa para quien no tuvo contacto con lo que designa, y cómo esa función puede cambiar.

Un experimento de 2021 muestra el fenómeno en miniatura. Cinco estudiantes universitarios aprendieron, por ensayo y error, qué figuras abstractas "combinaban" con cada una de tres fotografías de presidentes de Brasil (Bolsonaro, Lula y Fernando Henrique Cardoso). Antes y después del entrenamiento, evaluaron fotografías y figuras en una escala de adjetivos opuestos. Tres de los cinco evaluaron las figuras abstractas como neutras en el pretest. En el postest, las figuras pasaron a recibir evaluaciones semejantes a las de la fotografía a la que habían sido asociadas (Pereira & Moreira, 2021b).

Son cinco participantes, un diseño de pretest y postest y una escala. El capítulo trata el resultado como lo que es: una demostración de que el mecanismo existe en esas condiciones, y no una estimación de cuán común es. Lo que el experimento ofrece es un modelo explícito de cómo una evaluación llega a un estímulo sin pasar por el contacto con él. Las secciones 4.1 a 4.3 construyen ese modelo; la 4.4 reúne los estudios con estímulos de significado social; la 4.5 describe en detalle los dos trabajos con presidentes, el artículo de 2021b y los dos experimentos del libro de 2021a, con los estímulos y las pantallas que los propios autores usaron; las secciones 4.6 a 4.8 tratan de las relaciones de oposición, del cambio de actitud y de los límites de la evidencia.

Las figuras del capítulo siguen un código de colores que vale de la primera a la última. El azul marca las fotografías de los presidentes (conjunto A); el naranja y el amarillo, las dos series de figuras abstractas (B y C); el morado, las siglas de los partidos (D); el verde, las palabras *derecha*, *centro* e *izquierda* (E). En las flechas, el azul oscuro continuo es la relación enseñada, con acierto y error; el discontinuo verde oscuro, la simetría puesta a prueba; el discontinuo morado, la transitividad. Las fotografías, las figuras abstractas y las pantallas vienen de los trabajos de Pereira y Moreira (2021a, 2021b); los gráficos rehacen los números de las tablas de los autores; los esquemas hechos para enseñar a leer un gráfico llevan escrito, en la propia figura, que los valores son inventados.

4.1 La actitud como conducta verbal evaluativa

En la psicología social, el núcleo de la definición es la evaluación. Ajzen (2001) registra que hay acuerdo general en que la actitud representa una evaluación-síntesis de un objeto psicológico, en dimensiones como bueno–malo y agradable–desagradable. Bohner y Dickel (2011, p. 392) la definen como "una evaluación de un objeto de pensamiento" (traducción mía) y llaman objeto de actitud a todo lo que la persona puede tener en mente, de lo mundano a lo abstracto: cosas, personas, grupos e ideas. Albarracin y Shavitt (2018) dicen que estudiar actitudes es estudiar evaluaciones. La definición ya fue más amplia, con componentes cognitivo, afectivo y conductual, y se fue concentrando en el componente evaluativo (Mizael et al., 2016b). La disputa que queda es sobre el estatuto de la evaluación: Bohner y Dickel (2011) ordenan las definiciones en un eje que va de entidades estables, guardadas en la memoria, a juicios temporales, construidos en el momento.

El Análisis de la Conducta desplaza la pregunta: en lugar de qué estado revela la evaluación, qué contingencias la mantienen. Para Guerin (1994), las actitudes y las creencias son conducta verbal, y el moldeamiento por una comunidad verbal es parte esencial de la formación y del mantenimiento de ambas. Mizael et al. (2016b) resumen lo que cambia: para el área, las actitudes no solo están relacionadas con conductas, son conducta, abierta o encubierta, emitida o evocada según una historia de aprendizaje, y, como conducta verbal, pueden asumir funciones de tacto, de mando y otras, según las contingencias de la comunidad. Esto permite investigar la formación y el cambio de actitudes a partir de las relaciones con el ambiente, y no como algo causado por instancias "dentro de la persona" (p. 125). Cameoka y Moreira (2021) escriben que, entendidas como conducta verbal, independientemente de la topografía, las actitudes son conductas mantenidas por el reforzamiento social generalizado de la comunidad verbal a la que pertenece el individuo.

Ofrezco, para los fines de este libro, una definición en tres capas, cada una de las cuales añade una restricción a la anterior. La primera es la de la psicología social: la actitud es la evaluación de un objeto, con un valor que va de lo positivo a lo negativo. La segunda es analítico-conductual: esa evaluación es conducta, seleccionada y mantenida por consecuencias, y no un estado que causa la conducta. La tercera especifica el tipo: es conducta verbal, mantenida por el reforzamiento social que la comunidad verbal dispone, y por eso varía entre tacto, intraverbal y mando según la función que cumple para quien escucha (Capítulo 3). Las dos primeras capas dejan fuera al oyente. La tercera lo vuelve a poner, y por ella se puede leer una declaración de actitud como la leyó Guerin, por el efecto que tiene en quien escucha (Capítulo 2, sección 2.8).

CONCEPTO EN USO

Actitud (Guerin, 1994; Cameoka y Moreira, 2021)

La actitud es la respuesta de atribuir a un objeto (cosa, persona, grupo o idea) un valor que va de lo positivo a lo negativo, cuando esa respuesta es verbal: mantenida por las consecuencias que la comunidad verbal dispone para ella y, por eso, capaz de funcionar como tacto, intraverbal o mando. No es necesario haber tenido contacto con el objeto: el control puede estar en otras palabras e imágenes. La actitud difiere de la creencia porque el hablante puede sostenerla ante evidencia contraria (Capítulo 2, sección 2.8).

Dos precauciones de uso. La primera tiene que ver con la palabra *prejuicio*, que aquí sigue a Pereira y Moreira (2021b): una actitud en la que los atributos evaluativos son negativos. Mizael y de Rose (2017) la describen como un conjunto de actitudes negativas hacia individuos con determinadas características físicas, color o etnia. El rótulo es del investigador y se aplica a una relación (grupo–atributo negativo) que él eligió medir; el rótulo se mantiene con esa restricción, como se mantiene "discurso de odio" como rótulo externo (Capítulo 5). La segunda precaución tiene que ver con la medida. Lo que los experimentos registran es una marca en una escala o una elección en una pantalla, dirigida a un investigador. Esto es conducta verbal bajo control de quien pregunta (Guerin, 1994), y la sección 4.8 vuelve sobre lo que se puede y lo que no se puede concluir de ella.

4.2 Cómo una palabra adquiere función sin contacto con el referente

¿Cómo llega un nombre a tener valor para quien nunca tuvo contacto con lo que designa? Dixon et al. (2003) responden con un ejemplo elegido por ser simple. Una persona puede oír, en una ocasión, que los miembros de un grupo

nacional son grandes bebedores y, en otra, que los grandes bebedores son un poco obtusos (en el original, *dim-witted*). Nadie le enseñó que los miembros del grupo son obtusos, pero, según los autores, las dos relaciones enseñadas bastarían para derivar la relación entre el nombre del grupo y el atributo; un prejuicio infundado puede establecerse así por modos de discurso sobre el grupo, sin enseñanza directa. El artículo es conceptual: el ejemplo es una ilustración, no un dato.

Para saber si las relaciones enseñadas producen de hecho relaciones no enseñadas, hace falta un procedimiento en el que las enseñadas estén controladas y las derivadas, puestas a prueba. Es lo que ofrece el paradigma de equivalencia de estímulos, y a partir de él la literatura analítico-conductual pasó a entender la formación de actitudes como una red de relaciones arbitrarias entre clases de estímulos y atributos evaluativos (Mizael et al., 2016b). Bastan dos conceptos; los presento en recuadros porque el capítulo los usa de aquí en adelante.

CONCEPTO EN USO

Discriminación condicional e igualación a la muestra (Sidman y Tailby, 1982)

En un ensayo de igualación a la muestra (*matching-to-sample*, MTS), se presenta un estímulo muestra y dos o más estímulos de comparación, y la elección de uno de ellos es seguida de una consecuencia de acierto; cuál es el correcto depende de la muestra (Mizael et al., 2021b). El desempeño define una relación condicional entre estímulos: si A1, entonces B1; si A2, entonces B2 (Sidman & Tailby, 1982). La Figura 4.1 muestra un ensayo, con los estímulos de uno de los experimentos que el capítulo describe más adelante.

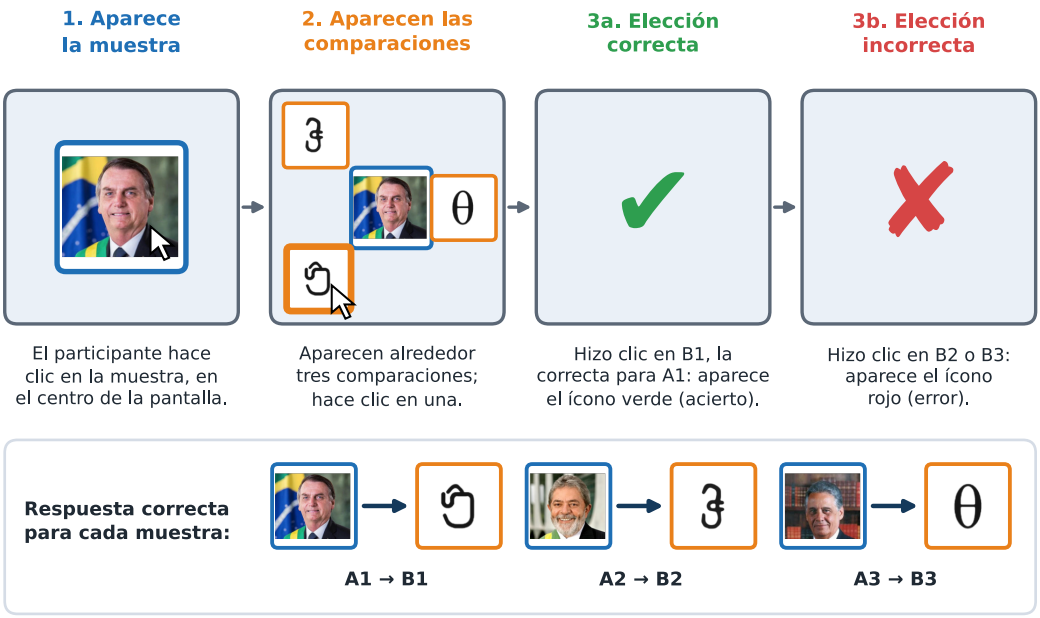
La figura sigue un ensayo del experimento en línea descrito en la sección 4.5. En el primer cuadro, el participante ve en el centro de la pantalla el estímulo muestra, aquí la fotografía de Bolsonaro, que los autores llaman A1, y hace clic en él. En el segundo, aparecen alrededor tres figuras abstractas, y él hace clic en una. Cuál de ellas es la correcta no está escrito en ningún lugar de la pantalla: lo definió el investigador, y el participante solo lo descubre por las consecuencias. Hacer clic en B1 ante A1 es seguido de un ícono verde (tercer cuadro), y hacer clic en B2 o en B3, de un ícono rojo (cuarto cuadro); después, la pantalla queda en blanco durante un segundo y comienza otro ensayo (Pereira & Moreira, 2021a). La franja de abajo resume lo que se enseña: ante la fotografía de Lula (A2), la elección correcta es B2; ante la de Fernando Henrique Cardoso (A3), es B3. Correcto, aquí, es solo lo que el procedimiento refuerza: nada en la tarea dice por qué esa figura combinaría con esa fotografía. En el

artículo, la pantalla era un poco diferente (la muestra arriba, las comparaciones abajo, una al lado de la otra, y un emoji feliz o triste como consecuencia), pero la lógica del ensayo es la misma (Pereira & Moreira, 2021b).

La condicionalidad, sin embargo, no dice que los estímulos tengan otra relación además de la que enseñó el procedimiento. Sidman y Tailby (1982) observan que la equivalencia, a diferencia de la condicionalidad, no puede definirse solo por la interacción del sujeto con el procedimiento que la estableció: exige pruebas, derivadas de las tres propiedades que la matemática elemental especifica como definición de la relación de equivalencia.

FIGURA 4.1

Un ensayo de igualación a la muestra, paso a paso



Nota. Montada con las fotografías y las figuras abstractas de los dos trabajos de Pereira y Moreira (2021a; 2021b) y con la disposición de la pantalla del experimento en línea (2021a), reproducida en la Figura 4.10. "A1→B1" se lee así: con A1 como muestra, elegir B1.

CONCEPTO EN USO

Clase de equivalencia (Sidman y Tailby, 1982)

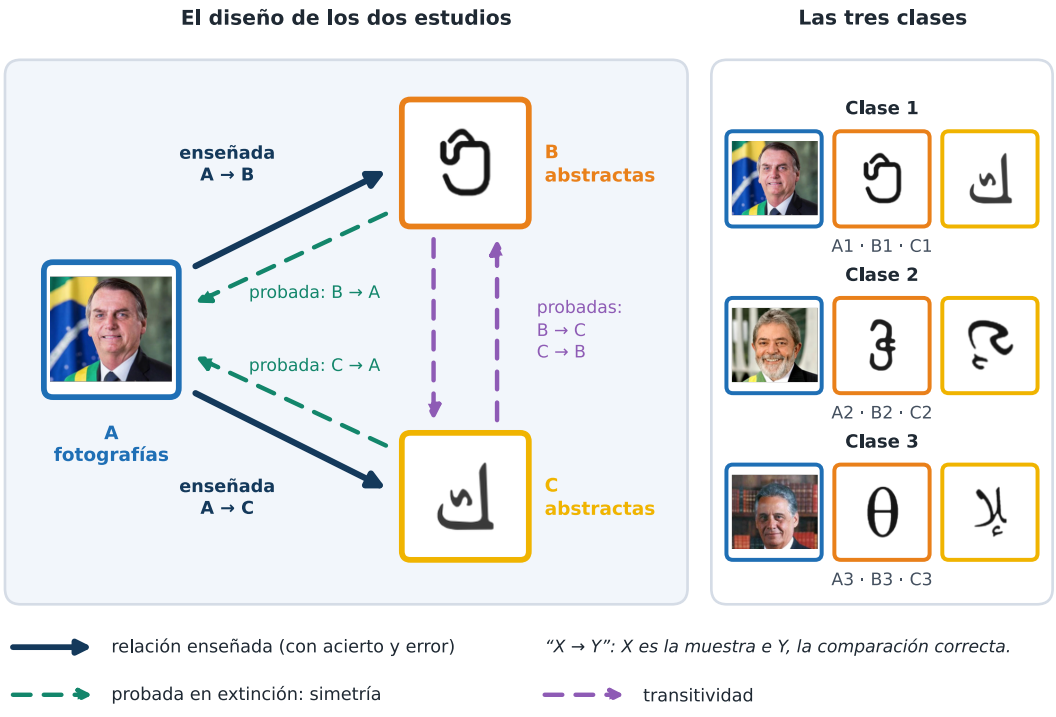
Un conjunto de estímulos forma una clase de equivalencia cuando las relaciones condicionales entre ellos tienen las tres propiedades de la relación de equivalencia. *Reflexividad*: cada estímulo se relaciona consigo mismo (A→A), lo que se pone a prueba con igualación por identidad generalizada. *Simetría*: enseñada A→B, emerge B→A sin entrenamiento adicional. *Transitividad*: enseñadas A→B y B→C, emerge A→C sin reforzamiento diferencial. La prueba que combina simetría y transitividad (C→A, después de A→B y B→C) es la prueba de equivalencia

(Sidman & Tailby, 1982; Mizael et al., 2016b). En el paradigma original, con tres conjuntos de tres estímulos, seis relaciones enseñadas y seis emergentes definen tres clases de tres miembros (Sidman & Tailby, 1982).

Los dos estudios con presidentes, descritos en la sección 4.5, usan el paradigma con una variación: en lugar de una cadena A→B→C, se enseñan dos relaciones a partir del mismo estímulo, A→B y A→C, donde A son fotografías de presidentes y B y C, formas abstractas (Figura 4.2). No se enseña ninguna relación entre B y C. Las relaciones de B hacia A, de C hacia A y entre B y C se ponen a prueba sin consecuencias, y, cuando las elecciones siguen la disposición esperada, se dice que las tres clases, una por fotografía, se formaron. Las relaciones de B hacia C y de C hacia B exigen combinar la vuelta de B hacia A, que es simetría, con la ida de A hacia C; en la terminología del recuadro anterior, son pruebas de equivalencia, aunque Pereira y Moreira (2021b) las llamen relaciones de transitividad.

FIGURA 4.2

Relaciones enseñadas, relaciones puestas a prueba y clases resultantes en los estudios con presidentes



Nota. Diseño de los estudios de Pereira y Moreira (2021a, 2021b), con las fotografías y las figuras abstractas de los propios trabajos (2021a; 2021b). En cada ensayo de entrenamiento, la muestra es una fotografía (A) y las comparaciones son las tres figuras de un mismo conjunto (B o C). Las flechas discontinuas indican relaciones que el procedimiento no refuerza: se ponen a prueba sin

consecuencias. “X→Y” se lee: X es la muestra e Y, la comparación correcta. La reflexividad no aparece en el diseño: suele suponerse, y no verificarse, en participantes verbales (Mizael et al., 2016b).

Dos precisiones evitan lecturas erróneas de lo que se llama clase. La primera es que la clase se define por pruebas, no por semejanza: los miembros no tienen atributos físicos semejantes y, después del entrenamiento, se vuelven sustituibles unos por otros (Cameoka & Moreira, 2021), en ciertos contextos (Pereira & Moreira, 2021b). La segunda es que sustituibles no quiere decir equivalentes en cualquier ocasión: tratamos la palabra “pastel” y el pastel como equivalentes y, aun así, no nos comemos la palabra (Mizael & de Rose, 2017). La clase es una relación bajo contexto, y así es como se usa el término a continuación.

El interés del paradigma está en lo que explica. De Rose et al. (1988) enuncian el problema: una conducta nueva, que ocurre sin historia de reforzamiento explícita y que no se explica por generalización primaria de estímulos. Para estudiarla, enseñaron dos cosas a adultos, a niños preescolares y a adultos con discapacidad intelectual. La primera fue una discriminación simple: con las formas A1 y A2 en la pantalla al mismo tiempo, elegir A1 era reforzado (A1 era el S+) y elegir A2, no (A2 era el S−). La segunda fue una discriminación condicional por igualación a la muestra: con A1 o A2 como muestra y las formas B1 y B2 como comparaciones, elegir B1 ante A1 y B2 ante A2 era reforzado, de modo que B1 y B2 eran correctos con la misma frecuencia. Después, B1 y B2 se presentaron juntos, sin muestra, y ninguna elección en esas sondas era reforzada. Por la historia de reforzamiento de B1 y B2, nada favorecía a uno de ellos, y aun así los participantes casi siempre eligieron B1, la forma ligada condicionalmente a A1, el S+ de la discriminación simple. Los autores sugieren que B1 y B2 adquirieron, en el entrenamiento condicional, funciones semejantes a las de A1 y A2. Es el primer paso de lo que la sección siguiente llama transferencia de función.

4.3 Transferencia de función

Lo que hace que la clase de equivalencia sea relevante para la actitud es la transferencia de función: la función que un estímulo adquiere por contacto directo aparece también en los demás miembros de la clase, sin contacto directo. Dougher et al. (1994) definen la transferencia de función como la adquisición de función de estímulo en virtud de la pertenencia a una clase de equivalencia. Dymond y Rehfeldt (2000) hablan de transformación de funciones cuando las funciones de un estímulo alteran las de otro de acuerdo con la relación derivada entre ellos, sin entrenamiento adicional. La elección del nombre tiene historia: la mayor parte de la investigación se relató como transfe-

rencia, y los autores defienden la terminología genérica de la teoría de los marcos relacionales (p. 239); Sidman, por su parte, objetó que "transferencia" sugiere un proceso hipotético innecesario, y Hayes y colegas, que el término describe mal lo que ocurre en relaciones que no son de equivalencia (Weinstein et al., 2008). Uso *transferencia de función* para lo que ocurre por equivalencia y *transformación* cuando la relación es otra (sección 4.6).

Los experimentos que sustentan esto usan estímulos arbitrarios y funciones fáciles de medir. Las funciones discriminativas se extienden por las relaciones condicionales, como en el experimento de De Rose et al. (1988) descrito en la sección anterior. Las funciones de consecuencia también. En Hayes et al. (1991), nueve estudiantes de psicología introductoria aprendieron, por igualación a la muestra con formas visuales arbitrarias, tres clases de tres miembros. Después, los estímulos de comparación recibieron función de reforzador condicionado o de castigador al ser emparejados con retroalimentación verbal en una tarea de clasificación. En ocho de los nueve participantes, la función entrenada se transfirió por las respectivas clases sin entrenamiento adicional.

CONCEPTO EN USO

Transferencia (o transformación) de función (Dougher et al., 1994; Dymond y Rehfeldt, 2000)

Dado un estímulo que adquirió una función por contacto directo (discriminativa, de consecuencia, elicitadora o evaluativa), la función se pone a prueba en los demás miembros de la clase sin entrenamiento adicional (Dymond & Rehfeldt, 2000). Que la función haya aparecido en los demás es un dato; que se haya "transferido" por la clase es interpretación, que hay que distinguir de explicaciones por contacto directo (sección 4.8).

Las funciones elicitadoras también. En Dougher et al. (1994), ocho universitarias aprendieron, por igualación a la muestra, dos clases de cuatro miembros. Un miembro de una clase se emparejó con descarga eléctrica, y un miembro de la otra clase se presentó sin descarga; a continuación, se presentaron todos los demás estímulos, y el condicionamiento se midió por la conductancia de la piel, una medida fisiológica de respuesta emocional. La respuesta apareció en los demás miembros de la clase del estímulo emparejado con descarga en seis de las ocho participantes. En el segundo experimento, después de condicionar a todos los miembros de una clase, uno de ellos se presentó en extinción, es decir, sin descarga, y los demás estímulos de la clase dejaron de elicitar la respuesta; cuando el estímulo extinguido se recondicionó, los otros miembros recuperaron la función elicitadora.

Un hallazgo de Hayes et al. (1991) importa para la lectura de estudios con estímulos socialmente cargados. En el Experimento 4, incluso estímulos nuevos de retroalimentación fueron tratados como consecuencias: la propia estructura de la tarea establece la retroalimentación como consecuencia. La dirección del efecto, sin embargo, dependió de la transferencia de funciones específicas por las clases. En otras palabras, el procedimiento ya trae funciones al interior del laboratorio, y es necesario separarlas de las que transfirió la clase.

El vínculo con las actitudes viene de estudios en los que la función transferida es evaluativa. En Bortoloti y de Rose (2009), universitarios relacionaron con figuras arbitrarias, por igualación a la muestra, rostros con expresiones emocionales y formaron tres clases, cada una con una expresión emocional y tres figuras arbitrarias. Los juicios sobre las figuras en un diferencial semántico, una escala de pares de adjetivos opuestos, resultaron semejantes a las evaluaciones de los rostros, y más aún cuando el entrenamiento usó igualación demorada, es decir, con un intervalo entre la muestra y las comparaciones. En un segundo experimento, las figuras a un nodo de distancia de los rostros fueron juzgadas semejantes a ellos, y las figuras a tres nodos, no. La distancia nodal, es decir, cuántos estímulos separan al miembro evaluado del que tiene la función, entra así como variable: la transferencia no es uniforme dentro de la clase. Weinstein et al. (2008) llevaron la idea al estigma. Cincuenta universitarios (45 analizados) aprendieron por igualación a la muestra a responder a formas arbitrarias como si fueran estímulos ligados a la obesidad o a la delgadez. Después, en una prueba de asociación implícita (IAT), que estima el sesgo por la rapidez con que la persona asocia categorías a palabras positivas o negativas, con formas semejantes a las entrenadas, apareció un efecto de sesgo pequeño (d = 0,31), frente a d = 0,96 en el IAT tradicional de obesidad. La d es una medida de tamaño del efecto, la diferencia entre medias en unidades de desviación estándar; por convención, 0,2 es pequeño y 0,8, grande. Los autores atribuyen la diferencia a la duración: unos 45 minutos de entrenamiento frente a toda una vida de historia.

Cómo se lee una escala de diferencial semántico

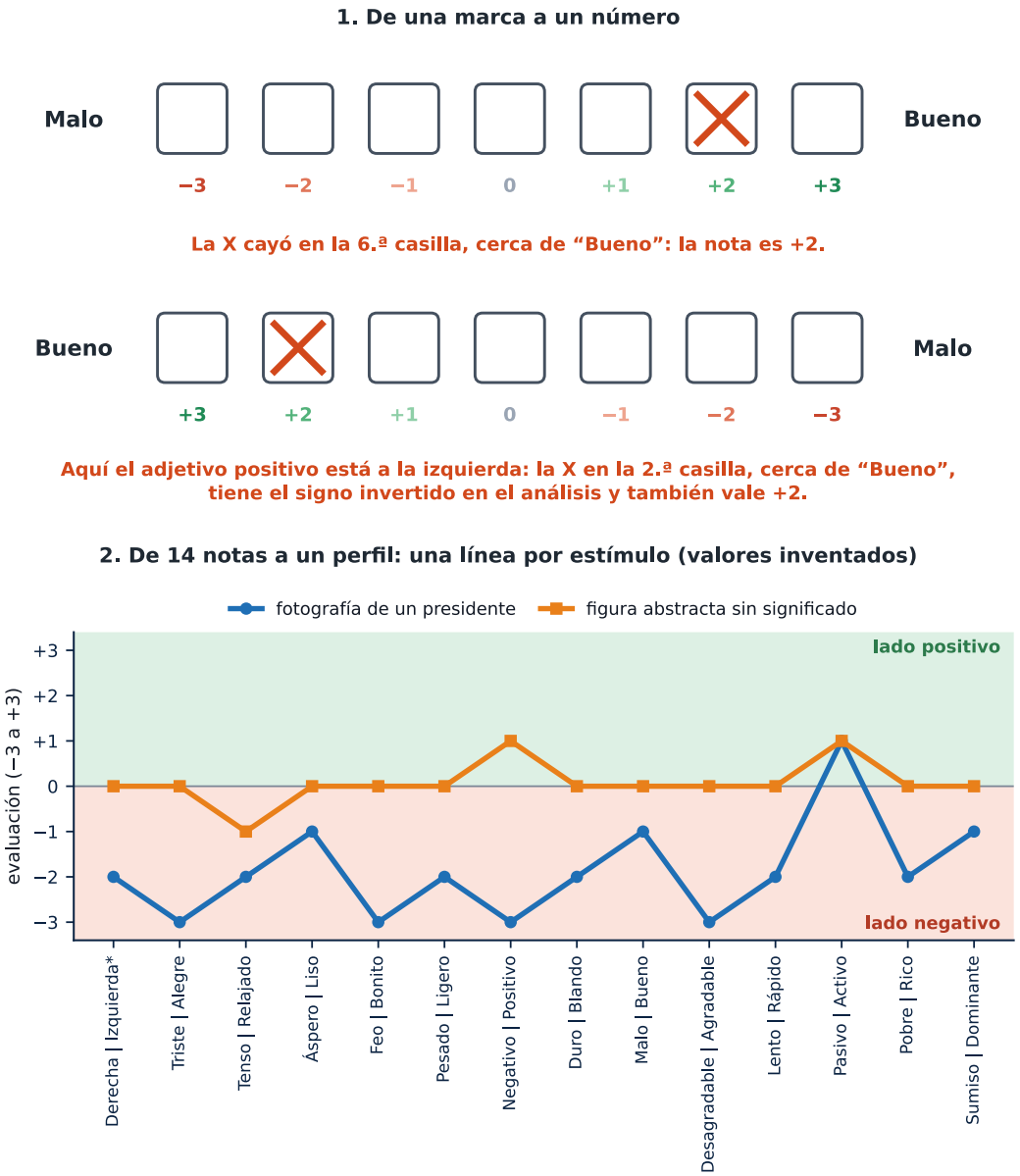
Los estudios con presidentes, que describe la sección 4.5, miden la evaluación con una escala de diferencial semántico, y saber leer sus gráficos es lo que permite seguir los resultados. La escala de Pereira y Moreira (2021b) tenía 14 pares de adjetivos bipolares, como malo y bueno, con siete casillas entre los adjetivos de cada par. Cada hoja traía un estímulo, y el participante marcaba, en cada par, una X en la casilla que le parecía describir mejor el estímulo: la más próxima a un adjetivo, si lo describía mucho; la del medio, si no lo descri-

bía ninguno de los dos. En la mitad de los pares el adjetivo positivo quedaba a la izquierda y en la otra mitad, a la derecha, de modo que marcar siempre del mismo lado no produce una evaluación uniforme.

La Figura 4.3 muestra, en la parte de arriba, cómo una marca se convierte en un número. Para el análisis, los autores dieron a cada casilla un valor de −3 a +3 e invirtieron el signo de los pares en los que el adjetivo positivo estaba a la izquierda, de modo que, en todos los pares, el adjetivo negativo quedó del lado de −3 y el positivo, del lado de +3. En la parte de abajo de la figura, las 14 notas de un estímulo, unidas por una línea, forman el perfil de evaluación de ese estímulo. Se lee cuánto se aparta la línea del cero, hacia el lado negativo o hacia el positivo, y comparar dos líneas es comparar dos evaluaciones. Los autores eligieron los gráficos de líneas porque facilitan ver la "forma de la evaluación" de cada estímulo y compararlas, aunque las notas de pares diferentes no sean una medida continua de una misma conducta.

FIGURA 4.3

Cómo se lee una escala de diferencial semántico: de una marca a un número y de 14 números a un perfil



** Derecha/Izquierda: el artículo convierte este par con la convención propia descrita en la sección 4.8.*

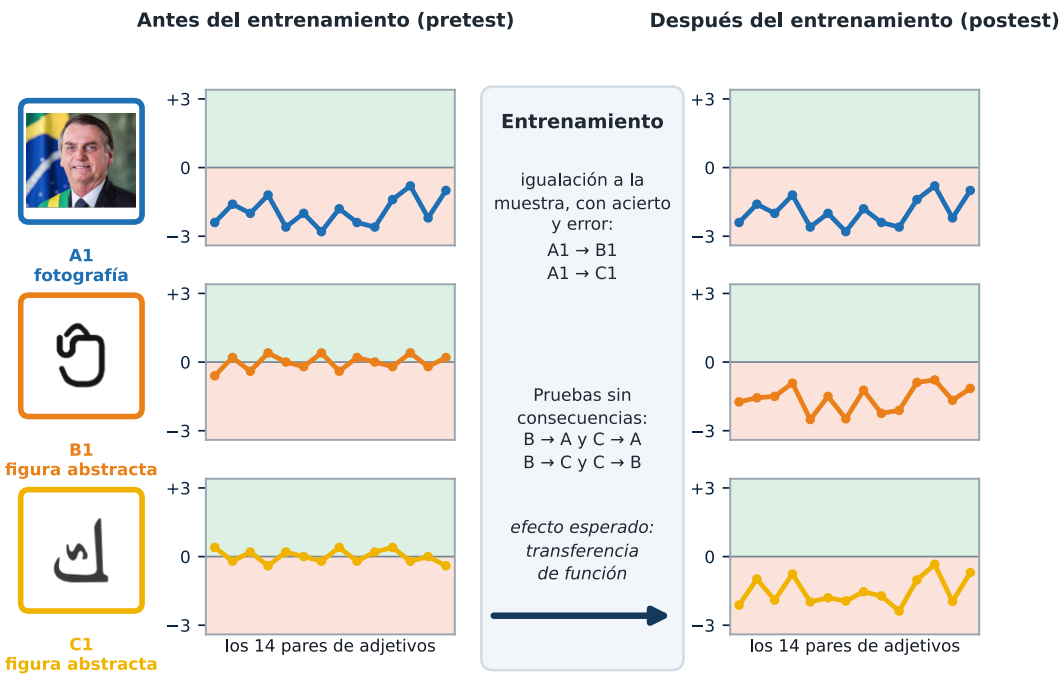
Nota. Esquema del autor. Los 14 pares son los de la escala de Pereira y Moreira (2021b), con el adjetivo negativo a la izquierda, como en el análisis; en el par Sumiso–Dominante, el artículo no dice qué adjetivo se trató como negativo, y la figura sigue el orden de la lista del artículo. Los valores de los dos perfiles son inventados, solo para enseñar a leer el gráfico. * El par Derecha–Izquierda sigue una convención propia del artículo (sección 4.8).

Cómo sería la transferencia de una evaluación

La Figura 4.4 junta las piezas en un esquema de lo que la transferencia de función quiere decir para una evaluación. A la izquierda, los tres estímulos de una clase, tomados del experimento de Pereira y Moreira: la fotografía de un presidente (A1) y dos figuras abstractas (B1 y C1). En el pretest, la fotografía tiene una evaluación propia, negativa en este ejemplo, y las figuras abstractas, que el participante nunca vio, quedan cerca del cero, la marca neutra, que, como recuerda la sección 4.5, no es prueba de ausencia de significado. En el centro está el entrenamiento, en el que A1 es muestra para B1 y para C1, con acierto y error, seguido de pruebas sin consecuencias, que verifican si las figuras remiten de vuelta a la fotografía y si se relacionan entre sí. A la derecha, lo que se espera ver si la función se transfirió: las dos figuras pasan a tener evaluaciones parecidas a las de la fotografía, que no cambió.

FIGURA 4.4

Transferencia de función de una evaluación por una clase de equivalencia



Valores inventados, solo para mostrar la idea. Cada punto es un par de adjetivos (−3 a +3); la franja rosa es el lado negativo y la verde, el positivo.

Nota. Esquema del autor, con valores inventados, solo para mostrar la idea; no es un dato de ningún participante. Los estímulos son los del experimento de Pereira y Moreira (2021b), cuyo entrenamiento y cuyas pruebas describe la sección 4.5. En los experimentos citados en esta sección con otras funciones, el evento es una descarga (Dougher et al., 1994) o retroalimentación verbal (Hayes et al., 1991), y los estímulos son formas arbitrarias.

DATO	En experimentos con estímulos arbitrarios, funciones discriminativas, de consecuencia, elicitadoras y evaluativas observadas en un miembro de una clase de equivalencia aparecieron en otros miembros sin entrenamiento directo (De Rose et al., 1988; Hayes et al., 1991; Dougher et al., 1994; Bortoloti & de Rose, 2009).
INTERPRETACIÓN	En esas condiciones, la pertenencia a la clase basta para que la función aparezca en miembros que no la tuvieron por contacto directo. La transferencia puede variar con parámetros del procedimiento y con la distancia nodal.
HIPÓTESIS	Que el mismo proceso opere sobre nombres, imágenes y símbolos de grupos y de políticos, en la historia de una persona fuera del laboratorio, es una extrapolación. Las secciones 4.4 y 4.5 examinan versiones de ella, con los límites allí descritos.

4.4 El prejuicio en los experimentos

El prejuicio entra en los experimentos de equivalencia por una diferencia de material. Con formas arbitrarias, el participante llega al laboratorio sin historia con los estímulos; con rostros, nombres, emblemas y palabras que designan grupos, no: llega con relaciones ya mantenidas por la comunidad verbal. Los estudios de esta sección se dividen, a grandes rasgos, en dos diseños. En uno, el entrenamiento enseña relaciones que no entran en conflicto con esa historia, y se pregunta si la función se transfiere a estímulos arbitrarios (Weinstein et al., 2008; Pereira & Moreira, 2021b). En el otro, el entrenamiento enseña relaciones que la contrarían, y se pregunta si la clase se forma y si la actitud cambia. Mizael et al. (2016a) dieron a este segundo diseño el nombre de *Conflicting Relations Paradigm* (Mizael et al., 2016b). La Tabla 4.1 reúne los estudios con estímulos de significado social cuyo texto íntegro se leyó; los dos trabajos con presidentes, que forman parte del primer diseño, se describen aparte, en la sección 4.5.

TABLA 4.1

Experimentos de equivalencia con estímulos de significado social: qué se hizo, qué se observó y qué no permiten decir

Estudio	Participantes y estímulos	Qué se observó	Salvedad
Weinstein et al. (2008)	50 universitarios (45 analizados); formas arbitrarias relacionadas, por igualación a la muestra, con palabras sobre obesidad y delgadez; medida por IAT	Sesgo en el IAT con formas semejantes después del entrenamiento (d = 0,31), frente a d = 0,96 en el IAT tradicional	Entrenamiento de unos 45 min; cinco excluidos por baja fluidez; el autoinforme no se correlacionó con el IAT tradicional
Dixon y Lemke (2007)	12 universitarios estadounidenses; fotografías de hombres de Oriente Medio, de hombres estadounidenses y de objetos; escala de 1 (malo) a 10 (bueno)	La mayoría invirtió el orden de las evaluaciones del pretest: los hombres de Oriente Medio pasaron a recibir evaluaciones más altas que los estadounidenses	Muestra pequeña; el sesgo se ligaba, probablemente, a los atentados de 2001, más intergrupal que racial (Cameoka & Moreira, 2021)
McGlinchey y Keenan (1997)	36 participantes en Irlanda del Norte (28 niños de 5 a 11 años, dos jóvenes de 12 y seis adultos); clases con la bandera británica, un estímulo arbitrario y una bailarina irlandesa	23 de los 28 niños respondieron por equivalencia (los otros cinco, no en al menos una prueba); los dos jóvenes de 12 años y los seis adultos, en ninguna (13 de 36, o 36%)	La edad confunde la duración de la historia con otras diferencias (salvedad del autor)
Haydu et al. (2015)	28 hombres de barras organizadas de tres clubes de São Paulo, residentes en Londrina (PR, estado de Paraná); emblemas de los clubes, pinturas abstractas y las palabras Bueno, Malo y Regular	Se formaron las relaciones de línea de base y de simetría; la transitividad y la equivalencia no emergieron; los errores, en su mayoría, en el sentido de la historia	Solo hombres, de una ciudad; palabras distribuidas a propósito para entrar en conflicto con la historia
Moxon et al. (1993)	19 universitarios (10 hombres y 9 mujeres); ocupaciones tradicionalmente masculinas, sílabas sin sentido y nombres femeninos; ocupaciones tradicionalmente femeninas como estímulos nuevos en la prueba	Respuestas de equivalencia en el 76% de las posibles en las mujeres y el 52% en los hombres; estímulo nuevo elegido en el 18% y el 42% de las ocasiones; generalización inconsistente, sin diferencia entre los grupos	N = 19; estereotipia de los participantes no medida (lectura del autor): la explicación por la historia es de los autores del estudio

Estudio	Participantes y estímulos	Qué se observó	Salvedad
de Carvalho y de Rose (2014)	4 niños de 7 a 10 años con sesgo negativo contra hombres negros, seleccionados entre 16	Uno revirtió las relaciones del pretest; en tres, persistieron	N = 4; los rostros pueden controlar el responder por la expresión y otras propiedades (Cameoka & Moreira, 2021)
Mizael et al. (2016a)	13 niños con sesgo negativo contra rostros negros; formas relacionadas con rostros negros y con un símbolo positivo; entrenamiento modificado	Los 13 formaron las clases; nueve mantuvieron las relaciones en una prueba más rigurosa; la diferencia entre las evaluaciones de rostros negros y blancos dejó de ser significativa	Cuatro de 13 no las mantuvieron en la prueba rigurosa (cálculo del autor)
Mizael et al. (2021b)	46 niños con sesgo negativo, en cuatro grupos (uno de control y tres con parámetros diferentes); seguimiento de seis semanas	33 formaron clases entre rostros negros y el símbolo positivo; 27 mantuvieron al menos una relación a las seis semanas; sesgo medio positivo en el postest para quienes las formaron y negativo para quienes no	Asociación, no causa; el grupo de control tuvo una reducción del sesgo y un mantenimiento menos pronunciados

Nota. Los números vienen de los artículos (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). El recuento 13 de 36 en McGlinchey y Keenan (1997) es del autor y coincide con el "36% de los casos" del texto del artículo.

Tres diseños de la tabla y de su entorno muestran cómo se arman estos estudios. En McGlinchey y Keenan (1997), seis adultos y treinta niños, católicos y protestantes de Irlanda del Norte, aprendieron cuatro discriminaciones condicionales: ante A1, un estímulo protestante (la bandera británica), elegir B1, una forma arbitraria; ante A2, un estímulo católico (la bailarina irlandesa), elegir B2; y, después, ante B1 elegir C1, un estímulo católico, y ante B2, C2, uno protestante. Una prueba de equivalencia verificó si los símbolos católico y protestante habían quedado relacionados por simetría y transitividad, y pruebas con estímulos nuevos verificaron si relaciones aprendidas socialmente competían con la emergencia de la equivalencia. Barnes et al. (1996) expusieron a 12 niños con discapacidad intelectual leve y a 12 sin discapacidad a dos procedimientos de equivalencia, uno con formas abstractas, que servía de control, y otro con las palabras "Slow" (lento) y "Able" (capaz) y con nombres personales, el del niño y un nombre ficticio, de modo que el entrenamiento ligaba el nombre del niño a "Able". Los niños con discapacidad tuvieron menos respuestas de equivalencia en esa prueba, es decir, fallaron en ligar su propio nombre a "Able", y los autores sugieren que relaciones preexperimentales en-

tre su nombre y el término "Slow" pueden haber perturbado la formación de las relaciones inducidas. En Merwin y Wilson (2005), 32 participantes respondieron antes a dos medidas de autoinforme, de sufrimiento y de autoestima, y se dividieron en grupos de sufrimiento y de autoestima altos o bajos; después hicieron dos tareas de igualación a la muestra, una que relacionaba estímulos autorreferentes (yo, mí) con palabras de evaluación positiva y otra que los relacionaba con palabras de evaluación negativa. Quienes informaban sufrimiento alto y autoestima baja se equivocaron más en las pruebas de equivalencia de la tarea con palabras positivas.

La tabla no muestra un resultado; muestra un patrón con tres partes. Primera: con estímulos arbitrarios, las funciones evaluativas aparecen después de entrenamientos cortos, con un tamaño pequeño (Weinstein et al., 2008; Bortoloti & de Rose, 2009). Segunda: cuando la clase enseñada contraría la historia social, la formación se interrumpe con frecuencia, y la interrupción acompaña la fuerza de la historia: en McGlinchey y Keenan (1997), los individuos con historias de aprendizaje social más extremas, contrarias al entrenamiento, fueron los menos propensos a responder por equivalencia. En Moxon et al. (1993), con ocupaciones y nombres femeninos, los hombres eligieron el estímulo nuevo más que las mujeres (el 42% y el 18% de las ocasiones), y los autores atribuyen la diferencia al mayor control, entre los hombres, de clases estereotípicas aprendidas socialmente; el estudio no midió esa historia. Barnes et al. (1996) y Merwin y Wilson (2005), descritos más arriba, apuntan en la misma dirección, con relaciones preexperimentales entre el propio nombre y el término *lento* y con el sufrimiento y la baja autoestima informados. Tercera: el mismo contenido social produce resultados opuestos según el procedimiento. Desde 1 de cada 4 niños con sesgo negativo contra hombres negros que revirtió las relaciones del pretest (de Carvalho & de Rose, 2014) hasta 13 de 13 que formaron las clases (Mizael et al., 2016a), con el mismo tipo de estímulo; Mizael et al. (2016b) atribuyen la diferencia, al menos en parte, a parámetros del entrenamiento, como ir del protocolo simple al complejo y reducir gradualmente la consecuenciación en la última fase. La dificultad, por lo tanto, no es solo de la historia: también es del diseño.

La Tabla 4.1 no agota lo que se hizo en Brasil con este diseño. El Instituto Walden4 reunió, en un libro de acceso libre, experimentos del mismo tipo con estímulos de significado social: *Preconceito: doze experimentos e um paradigma* (Moreira et al., 2022b). Después de un primer experimento sobre control de estímulos y medición de actitudes, que retoma la sección 4.8, el volumen trae experimentos de transferencia de función y de equivalencia con estímulos relacionados con la vestimenta de la mujer, con el género y el deporte, con el Candomblé, con el tatuaje en profesionales del área médica, con la obesidad,

con la resocialización de presos, con las personas mayores en el mercado de trabajo, con los usuarios de drogas y con los biotipos corporales (índice del volumen). Quedan fuera de la tabla porque este capítulo no los leyó en detalle; para quien quiera lo que la tabla no muestra, son el punto de partida más cercano.

DATO	Con estímulos de significado social, la clase pretendida no llegó a formarse en parte de los participantes (McGlinchey & Keenan, 1997: 13 de 36; Haydu et al., 2015: la transitividad y la equivalencia no emergieron; de Carvalho & de Rose, 2014: 3 de 4 no revirtieron las relaciones del pretest) y se formó en otros, con entrenamientos modificados (Mizael et al., 2016a: 13 de 13; Mizael et al., 2021b: 33 de 46).
INTERPRETACIÓN	La formación de clases que involucran estímulos socialmente cargados depende de la relación entre lo que se enseña y la historia del participante, y de parámetros del entrenamiento. Donde la clase enseñada contraría la historia, compite con relaciones ya mantenidas por la comunidad verbal.
HIPÓTESIS	La competencia crece con la duración y la fuerza de la historia de reforzamiento, lo que explicaría la diferencia entre niños pequeños y adultos en McGlinchey y Keenan (1997). Los estudios comparan edades (McGlinchey & Keenan, 1997) o sexos (Moxon et al., 1993); ninguno manipula la duración de la historia.

4.5 Actitudes hacia presidentes de Brasil

Dos trabajos de Pereira y Moreira (2021a, 2021b) son, entre los estudios de equivalencia verificados, los únicos con estímulos de política partidaria; los autores dicen no conocer otro (2021b). Uno es el artículo de *Acta Comportamentalia* (2021b), con cinco estudiantes de Psicología, escala en papel y entrenamiento en una computadora del aula. El otro es el libro (2021a), con dos experimentos de cinco estudiantes cada uno, hechos desde casa, por internet. Esta sección describe ambos con el detalle de un informe de método, uno por uno, en el orden en que el participante los vivió: qué vio, qué hizo y qué muestran los números. Las Figuras 4.5 a 4.13 acompañan el texto.

El artículo: pregunta, participantes y diseño

La pregunta del artículo era si un procedimiento de igualación a la muestra podría inducir la transferencia de función de estímulos con significado político, las fotografías de los presidentes, a estímulos sin significado, las figuras abstractas. Los autores esperaban que una respuesta afirmativa diera apoyo al

paradigma de equivalencia para el estudio de actitudes y añadiera evidencia con estímulos de significado social que no se ligan al prejuicio. La variable independiente fue el entrenamiento de discriminaciones condicionales; las variables dependientes fueron la formación de las clases, medida por el porcentaje de aciertos en las pruebas, y la transferencia de función, medida por la diferencia entre las evaluaciones de las figuras abstractas en las dos aplicaciones de la escala. El diseño es de sujeto único, con pretest y postest: cada participante sirve de comparación para sí mismo.

Participaron cinco estudiantes universitarios de 20 a 25 años, del quinto al séptimo semestre de Psicología, elegidos por conveniencia, sin contacto previo con la tarea; todos firmaron el consentimiento informado, y el proyecto fue aprobado por un comité de ética en investigación. La recolección se hizo en un aula de una institución de educación superior, con una computadora conectada a internet, una mesa y una silla; la igualación se ejecutaba en un software en línea elaborado por el segundo autor, que presentaba los estímulos y registraba las respuestas. Cada sesión duró de 80 a 100 minutos.

Los estímulos y las relaciones enseñadas

Los estímulos forman cinco conjuntos (Figura 4.5). El conjunto A reúne las fotografías de los tres presidentes: Bolsonaro (A1), Lula (A2) y Fernando Henrique Cardoso (A3). Los conjuntos B y C son dos series de tres figuras abstractas, sin significado conocido para los participantes; el D tiene las siglas de los partidos de los presidentes, PSL (Partido Social Liberal), PT (Partido dos Trabalhadores) y PSDB (Partido da Social Democracia Brasileira); y el E, las palabras *derecha*, *centro* e *izquierda*. Lo que el participante aprende es a relacionar A con B y A con C. Cada fotografía tiene, así, una clase pretendida, formada por ella y por una figura de cada serie: A1, B1 y C1; A2, B2 y C2; A3, B3 y C3. Las siglas y las palabras no se enseñaron: entran solo en la segunda prueba, para verificar si el participante ya las relacionaba con las fotografías y las figuras. Solo los conjuntos A, B, C y D se evaluaron en la escala, en un total de 12 estímulos.

FIGURA 4.5

Los estímulos del artículo, por conjunto y por clase



Nota. Fotografías y figuras abstractas reproducidas de Pereira y Moreira (2021a; 2021b). Las siglas y las palabras se compusieron para la figura. El artículo no dice qué palabra corresponde a qué columna (E1, E2, E3), y por eso el conjunto E ocupa las tres columnas.

Las siete etapas

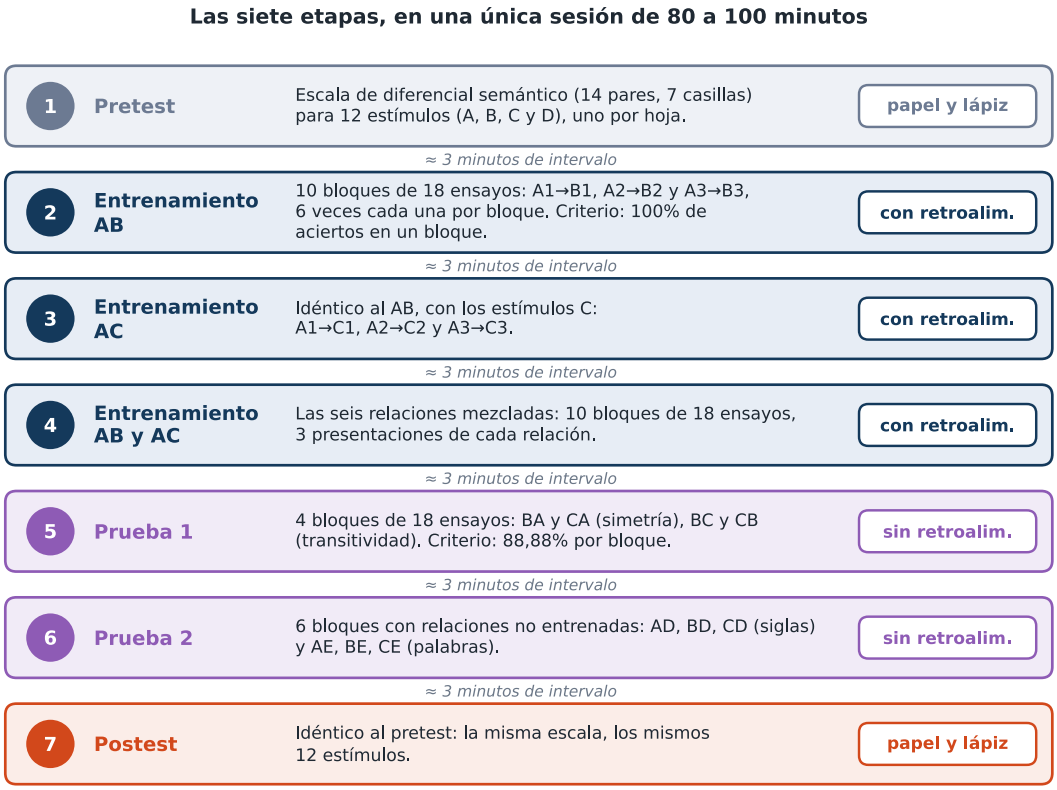
El procedimiento tuvo siete etapas, en una sesión única, con intervalos de unos tres minutos entre ellas, el tiempo que tardaba en hacerse en la computadora la programación de cada etapa (Figura 4.6). En el pretest, el investigador explicó oralmente cómo llenar la escala y entregó al participante un bloque de 12 hojas de papel A4, una para cada estímulo, evaluado una sola vez, en un orden fijo que empezaba por A1, C3, D2 y B1; en cada hoja, el participante marcaba una X en la escala de siete puntos en los 14 pares de adjetivos.

Los tres entrenamientos usaron igualación a la muestra simultánea. En cada ensayo, la muestra, una fotografía, aparecía en la parte central superior de la pantalla; el clic en ella hacía surgir tres comparaciones en la parte inferior, una al lado de la otra, y la muestra permanecía en la pantalla. Después del clic en una comparación, todos los estímulos salían de la pantalla y aparecía, durante un segundo, un emoji feliz, si la elección era la programada como correcta, o triste, si no lo era, seguido de dos segundos de pantalla en blanco. El

Entrenamiento AB tenía 10 bloques de 18 ensayos, con cada una de las tres relaciones (A1 ante B1, A2 ante B2 y A3 ante B3) presentada seis veces por bloque, en orden aleatorio; el criterio era el 100% de aciertos en un bloque, alcanzado en un máximo de 30 minutos, y quien no lo alcanzara repetiría el entrenamiento otro día. El Entrenamiento AC repitió el AB con las figuras C, y el Entrenamiento AB AC mezcló las seis relaciones, en 10 bloques de 18 ensayos, con tres presentaciones de cada una.

FIGURA 4.6

Las siete etapas del procedimiento del artículo



Nota. Esquema del autor, a partir del método de Pereira y Moreira (2021b). La sesión entera duró de 80 a 100 minutos.

En las pruebas, nadie le decía al participante si la elección era correcta o incorrecta: la primera prueba de emergencia de relaciones no tenía consecuencias diferenciales. Tenía cuatro bloques de 18 ensayos, con seis presentaciones de cada relación puesta a prueba, y verificaba, en este orden, las relaciones de simetría BA y CA (la figura como muestra y la fotografía como comparación) y las de transitividad BC y CB (una figura como muestra y las figuras del otro conjunto como comparaciones). El criterio era del 88,88% de aciertos en cada bloque, es decir, 16 aciertos en 18 ensayos, para afirmar que la relación había emergido. La segunda prueba, de seis bloques, verificó relaciones no entrena-

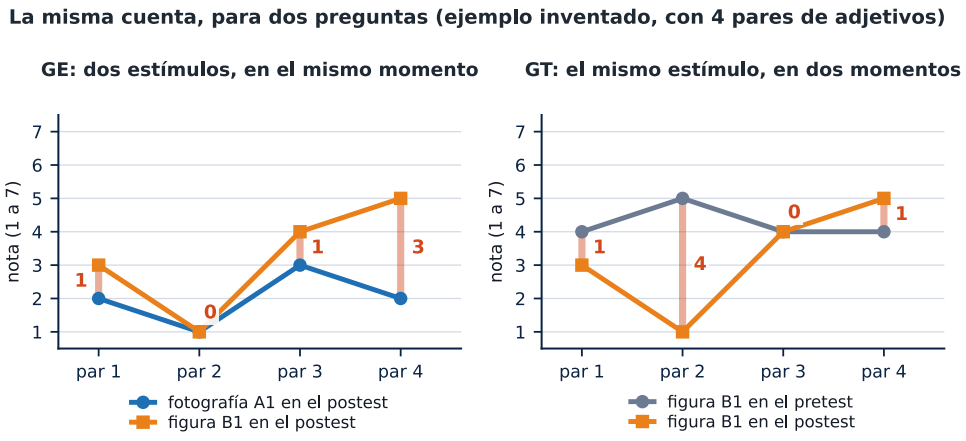
das con las siglas y las palabras: AD, BD y CD, por un lado, y AE, BE y CE, por el otro. Los autores advierten que el término emergencia no se aplica a las relaciones de las fotografías con las siglas y con las palabras (AD y AE), que probablemente ya existían antes del experimento. El postest fue idéntico al pretest.

Cómo se calculan el Grado de Equivalencia y el Grado de Transferencia

Para ir más allá de la inspección de los gráficos, los autores crearon dos medidas porcentuales. El Grado de Equivalencia (GE) mide la superposición entre las evaluaciones de dos estímulos diferentes de una misma clase en una misma prueba, por ejemplo A1 y B1 en el postest; un GE del 100% quiere decir evaluaciones idénticas en los 14 pares. El Grado de Transferencia (GT) mide la diferencia entre las evaluaciones de un mismo estímulo en pruebas diferentes, por ejemplo B1 en el pretest y B1 en el postest; un GT del 100% quiere decir que la evaluación cambió por completo, por ejemplo de −3 a +3 en los 14 pares. La cuenta es la misma para las dos (Figura 4.7). La escala se convierte de −3 a +3 en 1 a 7; se toma el valor absoluto de la diferencia entre las evaluaciones de los dos estímulos en cada par; se suman las diferencias de los 14 pares; como la mayor diferencia posible por par es 6, la mayor suma es 6 × 14 = 84; el porcentaje de esa suma con respecto a 84 es el GT, y el GE es 100 menos ese porcentaje.

FIGURA 4.7

Cómo se calculan el Grado de Equivalencia (GE) y el Grado de Transferencia (GT)



Con los 14 pares del estudio, el máximo es 6 × 14 = 84. GE = 100% quiere decir evaluaciones idénticas; GT = 100%, cambio total.

Nota. Ejemplo del autor, con cuatro pares de adjetivos y valores inventados, para mostrar la cuenta; el estudio usó 14 pares, y el máximo de la suma fue 84 (Pereira & Moreira, 2021b).

Con esto, los autores esperaban que los estímulos con significado preexperimental, como las fotografías, tuvieran un GT cercano al 0%, es decir, que sus evaluaciones no cambiaran; que los estímulos sin significado, como las figuras abstractas, tuvieran un GT cercano al 50%; y que esos mismos estímulos tuvieran un GE cercano al 100% con las fotografías de sus clases. Un GT del 50% equivale, en promedio, a mover tres puntos en cada par de la escala de seis puntos de amplitud, lo que corresponde, por ejemplo, a ir del centro de la escala a uno de los extremos; la cuenta es del autor de este libro, y no aparece en el artículo.

Lo que encontró el artículo

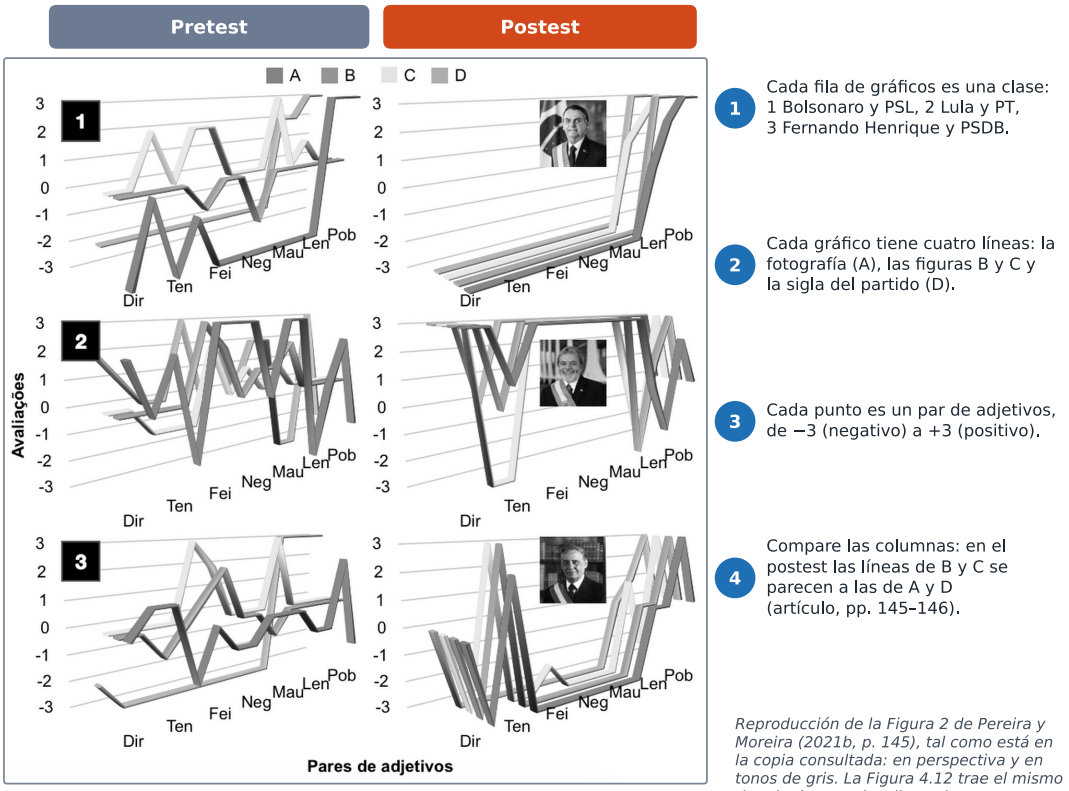
Todos los participantes alcanzaron el criterio del 100% de aciertos en los entrenamientos (AB, AC y AB AC), en la primera ejecución, y, en la primera prueba, todos alcanzaron el criterio del 88,88%, excepto P02 en la relación BA (72,22%); con esa salvedad, los autores concluyen que se formaron las tres clases programadas. En la segunda prueba, todos acertaron el 94,44% o más. En la Tabla 1 del artículo, casi todas las celdas son 100%: los demás valores son 94,44% (P01 en las relaciones CB y BE; P05 en las CB y BD) y, en P02, 72,22% en BA y 94,44% en CA.

La Figura 4.8 reproduce la Figura 2 del artículo, con los gráficos de P01. La columna de la izquierda es el pretest y la de la derecha, el postest; cada fila de gráficos es una clase; en cada gráfico hay cuatro líneas, la de la fotografía, las de las figuras B y C y la de la sigla del partido. Para P01, en la clase 1, la fotografía de Bolsonaro y la sigla PSL tuvieron evaluaciones predominantemente negativas en el pretest, con excepción de los tres últimos pares; la figura B1 quedó cerca de −1 y la C1, entre cero y tres. En el postest, la fotografía y la sigla mantuvieron las evaluaciones, y las de B1 y C1, diferentes de las del pretest, quedaron parecidas a las de la fotografía. En la clase 2, Lula y el PT fueron evaluados de modo predominantemente positivo en el pretest, y, en el postest, las figuras B2 y C2 pasaron a acompañar a la fotografía. La clase 3 muestra otro patrón: en el pretest, la fotografía de Fernando Henrique Cardoso varió entre neutra y negativa, y la sigla PSDB fue predominantemente negativa; en el postest, la sigla mantuvo la evaluación, y la fotografía y las figuras quedaron parecidas a ella, lo que los autores leen como transferencia de la sigla a los demás estímulos de la clase.

Los resultados de los otros cuatro participantes fueron semejantes a los de P01 en cuanto a la transferencia de función. Varía, entre ellos, el punto de partida de las figuras abstractas: P02, más que P01, les dio evaluaciones positivas y negativas ya en el pretest, lo que, en rigor, sugiere que las figuras tenían significado preexperimental para él, y P03, P04 y P05 les dieron nota cero, neutra.

FIGURA 4.8

Evaluaciones de P01 en el pretest y en el postest, en las tres clases



Reproducción de la Figura 2 de Pereira y Moreira (2021b, p. 145), tal como está en la copia consultada: en perspectiva y en tonos de gris. La Figura 4.12 trae el mismo tipo de dato en dos dimensiones.

Eje horizontal: los 14 pares de adjetivos; eje vertical: la evaluación, de −3 a +3.

Nota. Reproducida de Pereira y Moreira (2021b), tal como aparece en la copia consultada: en perspectiva y en tonos de gris, aunque el artículo dice que los gráficos se hicieron en color. Se añadieron las franjas de pretest y postest y la guía de lectura. Las figuras equivalentes de los demás participantes son las Figuras 3 a 6 del artículo. Los rótulos están en portugués, como en el estudio original: Avaliações (evaluaciones) y Pares de adjetivos (igual que en español); las abreviaturas de los pares vienen de los adjetivos en portugués.

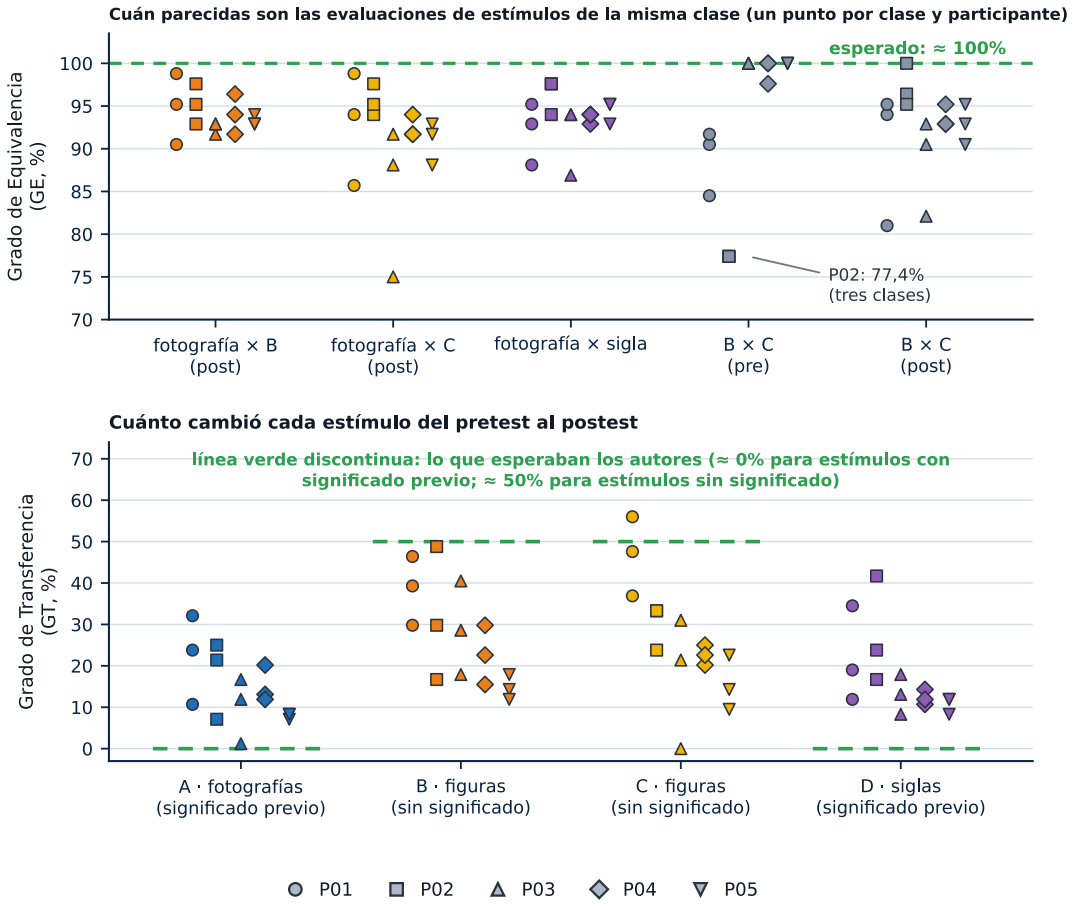
La Figura 4.9 reúne los números de la Tabla 2 del artículo, para los cinco participantes y las tres clases: cada columna tiene 15 puntos. En la parte de arriba, el GE en el postest entre cada fotografía y la figura B de su clase quedó por encima del 90% en todos los participantes y clases; entre la fotografía y la figura C, el patrón se repitió, con diferencias inferiores a 5 puntos con respecto al de B, excepto en la clase 3 de P03 (75,0%, frente al 92,9% en la figura B). El GE entre las figuras B y C, en el pretest, fue mayor del 90% en la mayoría de

los casos, excepto en P02, en quien quedó en el 77,4% en las tres clases. El GE, por sí solo, podría inducir a una lectura engañosa, si las evaluaciones de B en el pretest ya fueran parecidas a las de A, y por eso se calcularon también los GT.

En la parte de abajo de la figura, el GT de las figuras B fue del 11,9% (P05, B3) al 48,8% (P02, B1); el de las figuras C, en la tabla, va del 0,0% (P03, clase 3) al 56,0% (P01, clase 1), y solo ese valor supera el 50% que los autores esperaban para estímulos sin significado. Las fotografías y las siglas, que se esperaba que fueran estables, tuvieron un GT del 1,2% al 32,1% y del 8,3% al 41,7%, respectivamente. Para P01, por ejemplo, la evaluación de la fotografía de Lula varió un 23,8% y la de Fernando Henrique Cardoso, un 32,1%. Los autores admiten que no queda claro si esas variaciones en los estímulos con significado previo son un margen de error de medidas repetidas o transferencia de función de los demás estímulos a las fotografías, y especulan que la transferencia tal vez no sea un fenómeno unidireccional.

FIGURA 4.9

Grado de Equivalencia y Grado de Transferencia de los cinco participantes



Nota. Valores de la Tabla 2 de Pereira y Moreira (2021b), sin nuevos redondeos; cada punto es una clase de un participante. Colores: los de los conjuntos de estímulos (Figura 4.5); el gris marca las comparaciones entre B y C. Las líneas discontinuas verdes muestran lo que esperaban los autores.

El libro: dos experimentos en línea

El libro de Pereira y Moreira (2021a) rehace el estudio en otro formato. El Experimento 1 partió de la pregunta de si es posible transferir función de estímulos políticos a estímulos abstractos con una disminución progresiva de la retroalimentación, propuesta en un trabajo anterior de otros autores: todas las relaciones se entrenan juntas y la retroalimentación disminuye poco a poco, para reforzar de modo intermitente y obtener mayor resistencia a la extinción en la prueba; los autores dicen que hay solo un estudio anterior con estímulos políticos en ese paradigma. Participaron cinco estudiantes universitarios de 18 a 25 años, de carreras como Psicología, Publicidad, Letras, Odontología y Medicina, invitados por mensaje de WhatsApp en grupos de estudiantes y elegidos por conveniencia; la recolección solo empezó después de la aprobación del comité de ética y del consentimiento informado.

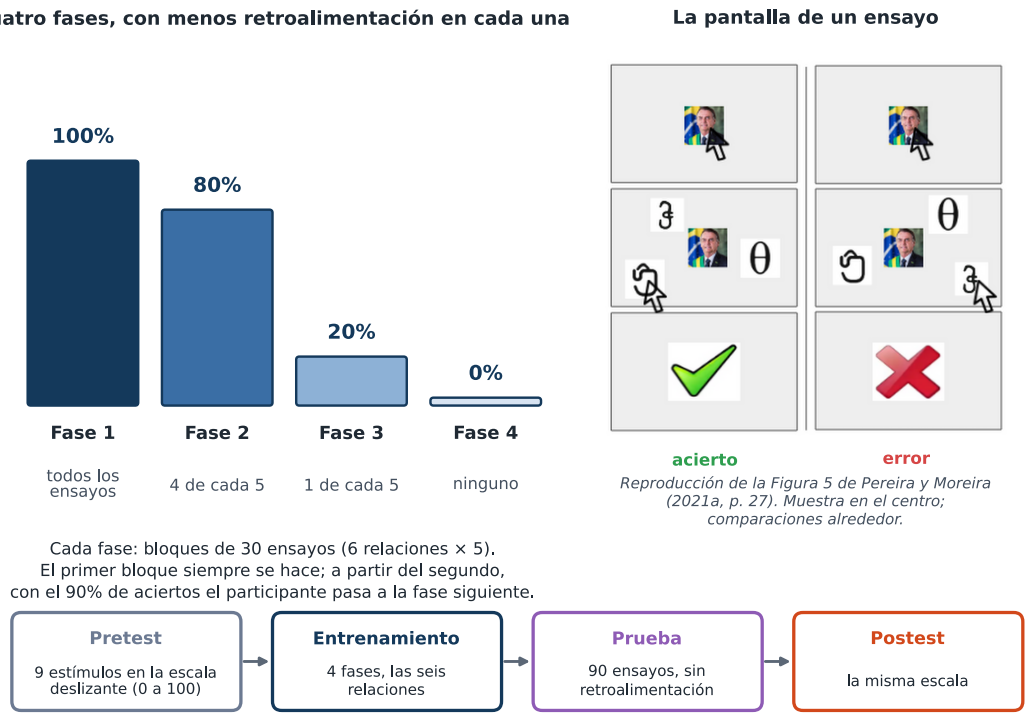
El experimento fue enteramente en línea. Cada participante estaba en su casa, solo en la habitación, en una computadora con internet de al menos 10 Mbps; quien tuviera menos sería excluido. Una plataforma presentaba los ensayos y la retroalimentación de modo automático y aleatorio y guardaba los resultados. Las instrucciones venían en videos estandarizados, con una asistente virtual animada, en texto y en imágenes; durante cada actividad, el investigador apagaba el audio y la cámara, y, entre las etapas, el participante podía hacer una pausa de hasta cinco minutos. La escala tenía 13 pares de adjetivos bipolares y se respondía en una barra deslizante, de 0, más cerca del adjetivo negativo, a 100, más cerca del positivo, en lugar de los siete puntos usuales. En el pretest, cada participante evaluó nueve estímulos, en orden aleatorio: tres fotografías de presidentes (A1, Bolsonaro; A2, Lula; A3, Fernando Henrique Cardoso) y seis figuras abstractas (dos conjuntos de tres); el botón para avanzar solo se habilitaba después de 15 segundos y una vez evaluados todos los pares.

El entrenamiento tenía cuatro fases, con cada vez menos retroalimentación (Figura 4.10). En cada ensayo, la muestra aparecía en el centro de la pantalla, y, después del clic en ella, tres comparaciones aparecían alrededor; el acierto era seguido de un ícono verde con una “v” y el error, de un ícono rojo con una “x”, y luego venía un segundo de pantalla en blanco. En la primera fase, la retroalimentación venía en todos los ensayos (100%); en la segunda, en cuatro de cada cinco (80%); en la tercera, en uno de cada cinco (20%); en la cuarta, en ninguno (0%). En cada fase, cada una de las seis relaciones se presentaba cinco veces, en bloques de 30 ensayos en orden aleatorio. El primer bloque siem-

pre se hacía; a partir del segundo, con al menos el 90% de aciertos el participante pasaba a la fase siguiente, y, si no alcanzaba el criterio en 30 minutos, el sistema terminaba la tarea y el participante quedaba descalificado. Por último, la prueba de simetría y transitividad era un único bloque de 90 ensayos aleatorizados, sin retroalimentación, con cinco ensayos por relación: las seis de la línea de base, las de simetría (B y C como muestra, A como comparación) y las de transitividad (BC y CB).

FIGURA 4.10

El experimento en línea del libro: cuatro fases de entrenamiento y la pantalla de cada ensayo



Nota. La pantalla es una reproducción de la Figura 5 de Pereira y Moreira (2021a); el resto es esquema del autor, a partir del método de la obra.

Lo que encontró el libro

Con la retroalimentación en descenso, casi todos los errores del entrenamiento ocurrieron cuando era total. Sumando las Tablas 1 a 5, del Experimento 1, y 27 a 31, del Experimento 2, hubo 115 errores, de los cuales 104 en la fase del 100% de retroalimentación, cinco en la del 80%, dos en la del 20% y cuatro en la del 0% (la suma es del autor de este libro; Figura 4.11). El mayor número de errores fue el de P5, con 36 errores en los 120 ensayos que hizo en la primera fase, y el de P1, con 20 errores en 90 ensayos; en las otras fases, todos hicieron 60 ensayos y se equivocaron como máximo dos veces.

En la prueba del Experimento 1, los errores fueron raros: P1 tuvo uno (una relación de transitividad, c1b1), P2 ninguno, P3 dos (ambos en b1c1), P4 dos (b2c2 y c1b1) y P5 ninguno, de modo que todos acertaron más del 90% de las relaciones en cada etapa de la prueba. Para saber si la evaluación se transfirió, los autores calcularon la semejanza entre las evaluaciones de dos estímulos de una clase como 100 menos el valor absoluto de la diferencia entre las dos notas; cuanto más cerca de 100, más parecidas las evaluaciones, y la diferencia entre la semejanza del postest y la del pretest, positiva si la semejanza aumentó, muestra lo que hizo el entrenamiento. Las relaciones se llaman ab (fotografía y primer conjunto de figuras), ac (fotografía y segundo conjunto) y bc (entre los dos conjuntos de figuras).

FIGURA 4.11

Errores de cada participante en cada fase de entrenamiento

Errores en cada fase de entrenamiento (porcentaje de ensayos con retroalimentación)

	100%	80%	20%	0%	
P1	20	0	0	1	Experimento 1 fotografías y figuras abstractas
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 rostros neutros y figuras abstractas
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	

En la fase del 100%, P1 hizo 90 ensayos, P5 hizo 120 y los demás 60; en las otras fases, 60 cada uno.

Nota. Números de las Tablas 1 a 5 y 27 a 31 de Pereira y Moreira (2021a); los totales por fase son del autor de este libro.

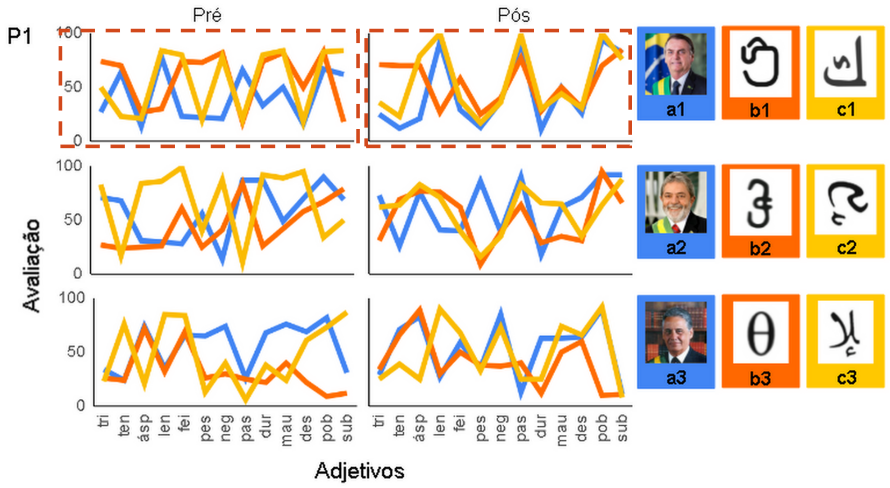
La Figura 4.12 reproduce la Figura 11 del libro, con las evaluaciones de P1 en las tres clases, en el pretest y en el postest. Las líneas azules son las notas del estímulo a (la fotografía); las naranjas, las del b; las amarillas, las del c, los mismos colores de los conjuntos de este capítulo; los 13 pares de adjetivos están en el eje horizontal, y las notas, de 0 a 100, en el vertical. Los autores describen que, para P1, las evaluaciones de los estímulos abstractos quedaron más parecidas a las de la muestra en la relación a1b1c1 en el postest que en el pretest; es lo que se ve en los dos recuadros discontinuos, en los que las tres

líneas se separan en el pretest y siguen más juntas en el postest. Como los pares de adjetivos tenían polos alternados, una limitación que registran los autores, lo que se compara en la figura es la distancia entre las líneas, y no la altura de cada una.

En la Tabla 26, el aumento medio de la semejanza fue de 9,4 puntos porcentuales en la relación ab, 14,1 en la ac y 6,0 en la bc (Figura 4.13). Los valores por participante varían: las medias de la relación ac quedaron entre 8,8 y 20,9, las de la ab entre 4,8 y 12,8, y la de la bc fue negativa en un participante, P5 (−0,9), es decir, la semejanza entre las dos series de figuras disminuyó un poco para él. Los autores observan también que, con excepción de P3, las evaluaciones de las fotografías cambiaron menos del pretest al postest que las de las figuras abstractas, como esperaban, porque los presidentes ya tenían un concepto previo; para P3, la semejanza de dos relaciones ya existía en el pretest, y por eso dicen que, solo por los gráficos, es difícil afirmar la transferencia de función.

FIGURA 4.12

Evaluaciones de P1 en el pretest y en el postest, en las tres clases (Experimento 1 del libro)



1. Cada línea de color es un estímulo: azul, la fotografía (a); naranja, la figura b; amarilla, la figura c.
2. Cada punto es un par de adjetivos (13 pares); la altura es la nota, de 0 a 100.
3. Compare los dos recuadros discontinuos (clase 1, Bolsonaro): después del entrenamiento, las tres líneas quedan más próximas (libro, p. 48).

Reproducción de la Figura 11 de Pereira y Moreira (2021a, p. 48).

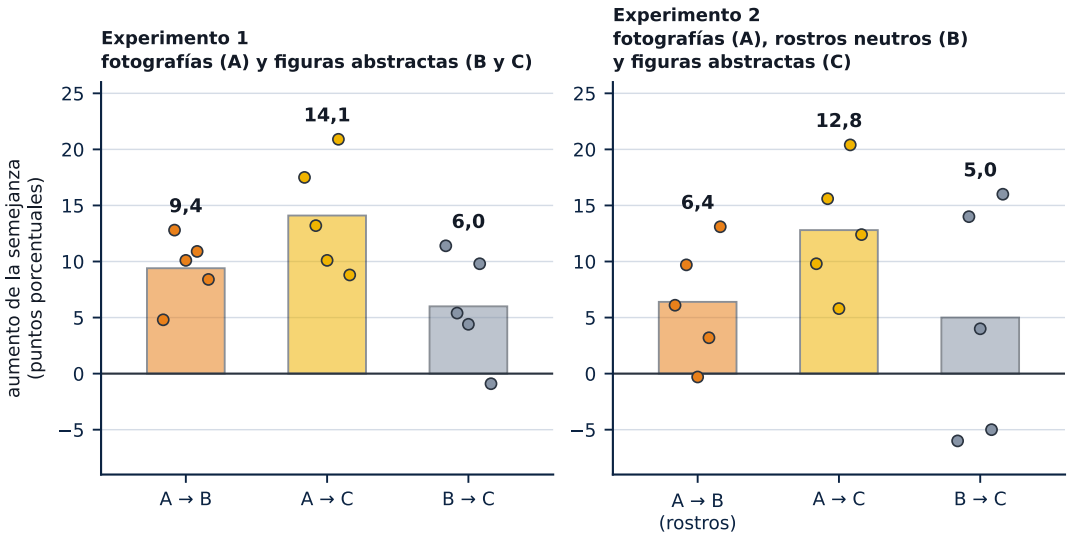
Nota. Reproducida de Pereira y Moreira (2021a), con el recuadro discontinuo y la guía de lectura añadidos. Los rótulos están en portugués, como en el estudio original: Pré/Pós (pretest/postest), Avaliação (evaluación) y Adjetivos (igual que en español); las abreviaturas vienen de los adjetivos en portugués.

El Experimento 2 parte de una hipótesis: cuanto mayor es la carga afectiva preexperimental de los estímulos, mayor es la dificultad para formar clases y para transferir función. Los autores la ilustran con un estudio sobre celos, en el que siete de 17 universitarios no formaron clases, y se proponen verificar la formación de clases y la transferencia con rostros de personas comunes, evaluados como neutros, e investigar si la transferencia es más fácil para estímulos abstractos que para figuras humanas. Los cinco participantes tenían un perfil y un reclutamiento idénticos a los del primer experimento. El conjunto B estuvo compuesto por rostros de personas desconocidas, de raza blanca, elegidos por haber sido los evaluados como más próximos a neutros en Cameoka y Moreira (2021), en una escala de 10 puntos (B1 y B2 con una media de 5,5 y B3 con 5,4); el conjunto C eran figuras abstractas, y el procedimiento fue idéntico al del Experimento 1.

En la prueba del Experimento 2, P6 tuvo seis errores en 90 ensayos (uno de identidad, uno de simetría y cuatro de transitividad, un 93,33% de aciertos en total), P7 y P10 ninguno, P8 uno y P9 uno. En la Tabla 52, el aumento medio de la semejanza fue de 6,4 puntos porcentuales en la relación ab, entre las fotografías y los rostros; de 12,8 en la ac, entre las fotografías y las figuras abstractas; y de 5,0 en la bc. El aumento en la ac fue el doble del de la ab. Por participante, P8 tuvo una media de −0,3 en la ab y de 5,8 en la ac, y, en su gráfico, las evaluaciones de comparación quedaron neutras también en el postest; en P9, la transferencia fue sobre todo de la fotografía de Lula a las figuras de la clase 2, menos en las clases 1 y 3, y no fue identificable para los rostros.

FIGURA 4.13

Aumento de la semejanza entre las evaluaciones, del pretest al postest, en los dos experimentos



Barra: media de los cinco participantes (total de las tablas). Puntos: cada participante. Por encima de cero, las evaluaciones quedaron más parecidas después del entrenamiento.

Nota. Valores de las Tablas 26 y 52 de Pereira y Moreira (2021a): barras, el total de los cinco participantes; puntos, la media de cada uno. Colores: los de los conjuntos de estímulos, con el gris para la relación entre B y C.

Para los autores, el aumento mayor en la relación ac que en la ab, de un 50% en el primer experimento y de un 100% en el segundo, indica que es más fácil cambiar la actitud hacia símbolos abstractos que hacia rostros neutros. Proponen que los rostros, aun elegidos por ser neutros, tienen características, como ser de un hombre, blanco, con barba y cabello, que les dan una carga preexperimental mayor que la de los estímulos abstractos, lo que puede haber dificultado la transferencia. Las limitaciones que registran son la recolección en línea, dependiente de la velocidad de internet y hecha en ambientes domésticos, y los pares de adjetivos con polos alternados; cada experimento se hizo en una sesión de una hora a una hora y media.

Lo que los dos trabajos permiten decir

Dos hallazgos del artículo importan para el argumento del capítulo. El primero es que las siglas de los partidos y las palabras *derecha*, *izquierda* y *centro* se relacionaron correctamente con las figuras abstractas en las pruebas, aunque no se había programado ninguna discriminación condicional con ellas; para los autores, esto sugiere que ya formaban parte, antes del experimento, de una clase con las fotografías de los presidentes. En el pretest, los participantes evaluaron sistemáticamente las fotografías de Bolsonaro, Lula y Fernando Henrique Cardoso como, respectivamente, derecha, izquierda y centro. La lectura que el capítulo hace de esto, a partir de la inferencia de los autores, es que el rótulo ideológico forma parte, para esos participantes, de la clase con la que el nombre del político ya llega al experimento. El segundo hallazgo es una precaución de los autores: las evaluaciones neutras, en torno al centro de la escala, no son ausencia de significado o de actitud, y P02 dio evaluaciones positivas y negativas a las figuras abstractas en el pretest, lo que sugiere un significado preexperimental para ellas.

Los autores también señalan lo que falta. Proponen replicar el procedimiento con otras estructuras de entrenamiento, series lineales en lugar de uno-a-muchos, para evaluar el efecto de la distancia nodal, ya que la semejanza esperada entre B y C vale en parte porque la distancia de A a B y de A a C era la misma. Y sugieren replicar el trabajo con fotografías de personas anónimas en lugar de las figuras abstractas, lo que demostraría la formación de actitudes hacia personas con las que el participante nunca interactuó, un fenómeno directamente ligado a prejuicios, estereotipos y discursos de odio. Ese es, a grandes rasgos, el diseño del Experimento 2 del libro, con rostros en lugar de figuras.

DATO	En cinco estudiantes, después de aprender relaciones entre fotografías de tres presidentes y figuras abstractas, las evaluaciones de las figuras abstractas en el diferencial semántico quedaron más próximas a las de la fotografía correspondiente; el Grado de Equivalencia entre fotografía y figura B superó el 90% en los cinco participantes y en las tres clases (Pereira & Moreira, 2021b). En el libro de los mismos autores, con diez participantes en dos experimentos, el aumento de la semejanza fue menor para rostros neutros que para figuras abstractas (Pereira & Moreira, 2021a).
INTERPRETACIÓN	La evaluación que cada participante ya hacía de cada político pasó a emitirse también ante estímulos sin historia con él, después de una sesión de alrededor de una hora y media. El diseño es de pretest y postest con sujeto único, sin condición de control sin entrenamiento; los autores no saben decir si parte de las diferencias en estímulos con significado previo es variación de medidas repetidas.
HIPÓTESIS	Que la transferencia sea menor para estímulos de mayor carga afectiva previa, como sugieren los autores de ese libro, es compatible con Haydu et al. (2015) y con Merwin y Wilson (2005), pero no se manipuló: los rostros y las figuras abstractas difieren en varias dimensiones. Que el rótulo derecha–izquierda sea parte de la clase con la que cada nombre llega al elector es un indicio de este estudio, en cinco estudiantes de Psicología, no un dato sobre electores. El Capítulo 7 (sección 7.6) retoma la idea para nombres y números de candidatos.

4.6 “Nosotros” y “ellos”: relaciones de oposición y de comparación

La equivalencia es una relación de coordinación, pero no es la única que derivan los humanos. Dixon et al. (2003) recuerdan que los patrones confiables de responder relacional, cuando quedan bajo control contextual de claves del ambiente (como las palabras “más que”), se llaman marcos relacionales, y citan la igualdad, la oposición, la diferencia, más que y menos que, antes y después. Weinstein et al. (2008) señalan lo mismo: la equivalencia no es la única relación funcional posible en la transformación de funciones. La diferencia importa porque, en una relación que no es de equivalencia, los estímulos no pasan a compartir funciones equivalentes; por eso los autores prefieren “transformación” a “transferencia”, que permite una descripción más amplia.

CONCEPTO EN USO

Relaciones derivadas de oposición y de comparación (Dixon et al., 2003)

Un marco relacional es un patrón de responder relacional bajo control contextual de claves del ambiente (por ejemplo, "es lo opuesto de", "es mayor que"). En la *oposición*, dos estímulos son lo opuesto de un tercero: si A es lo opuesto de Y y A es lo opuesto de Z, entonces Y es lo mismo que Z (Dixon et al., 2003; ejemplo ilustrativo de los autores, sin datos). En la *comparación* (más que, menos que), los estímulos relacionados no quedan con las mismas funciones: la relación define cómo se transforma la función, y no solo si se comparte (Weinstein et al., 2008).

Los estudios de equivalencia con estímulos sociales traen indicios de que la oposición opera. En McGlinchey y Keenan (1997), los participantes que no respondieron por equivalencia fueron entrevistados: eligieron estímulos nuevos porque los que formarían la relación esperada eran vistos como opuestos y, por eso, no combinaban. Algunos argumentaron que la bailarina irlandesa y la bandera británica no tenían nada en común y que, por lo tanto, la otra comparación "debía ser" la combinación. Los autores plantean que esos participantes pueden haber respondido bajo un marco de oposición, y relatan que un estudio anterior, de Moxon et al. (1993), intentó entrenar clases con estímulos vistos estereotípicamente como opuestos, y varios sujetos no respondieron por equivalencia; la descripción coincide con el artículo de Moxon et al. El dato es modesto: algunas justificaciones verbales, en entrevistas, después de la prueba.

Es un buen punto para leer "nosotros" y "ellos". En Haydu et al. (2015), las palabras se distribuyeron para entrar en conflicto con la historia de cada barra: el emblema del club apoyado con *malo* y el del principal rival con *bueno*; los errores más frecuentes fueron, de nuevo, elegir lo que la historia prescribía. Que la historia de una barra sea una historia de oposición entre "nosotros" y "ellos" es coherente con el diseño de los autores, que eligieron esa población por las rivalidades declaradas entre las barras.

■ DATO	En McGlinchey y Keenan (1997), participantes que no respondieron por equivalencia justificaron, en una entrevista, que los estímulos de la relación esperada eran vistos como opuestos.

INTERPRETACIÓN	Cuando la clase enseñada pone lado a lado estímulos que la comunidad verbal trata como opuestos, parte de los participantes responde según la oposición, y no según la coordinación que pide la tarea.
HIPÓTESIS	En el discurso político, la distinción "nosotros"–"ellos" funcionaría como una relación de oposición: atribuir un atributo aversivo a "ellos" transformaría la función de "nosotros" en el sentido inverso, y dos grupos opuestos a un mismo "ellos" pasarían a ser tratados como equivalentes entre sí (regla del ejemplo de Dixon et al., 2003). No hay datos, en este capítulo, sobre discurso; la hipótesis necesita declaraciones fechadas, como las de la ficha del Capítulo 3, y una prueba con el procedimiento de oposición.

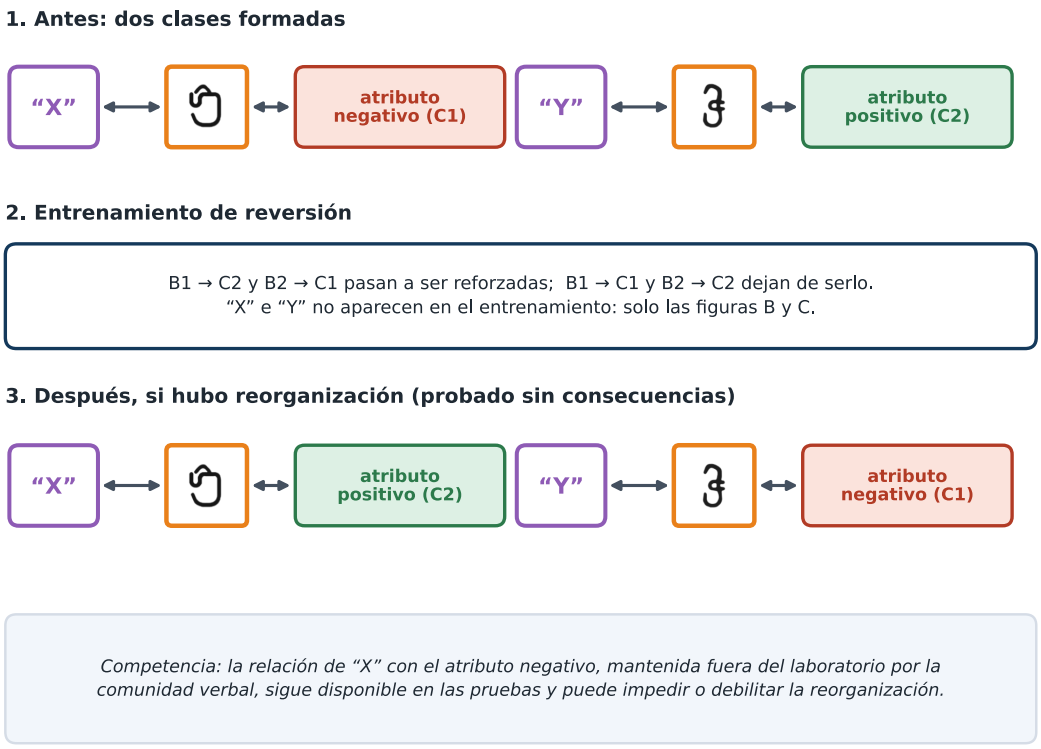
Dos observaciones de Dixon et al. (2003) ayudan a anticipar el Capítulo 12, y valen, como todo en ellos, como argumento sin datos. Instrucciones del tipo "no crea que todos los miembros del grupo X son Y" pueden aumentar la saliencia del grupo como concepto categórico y añadir atributos a la categoría. Y, en el análisis de los autores, el prejuicio y los actos de terror implican la transformación derivada de las funciones de individuos o grupos a partir de un contacto a menudo indirecto con pocos miembros del grupo. La sección 5.4 usa esta lectura para el rótulo de grupo.

4.7 Cambio de actitud: reorganización de las clases y cómics

En términos del paradigma, cambiar una actitud es reorganizar clases. Cameoka y Moreira (2021) describen la reversión de clases de equivalencia como un tipo de reorganización en el que una redisposición de las contingencias en la línea de base produce cambios en el responder coherentes con esa redisposición. La Figura 4.14 esquematiza el procedimiento con clases formadas en el laboratorio; en los estudios con rostros, la clase que hay que deshacer es la preexperimental, y el entrenamiento establece una clase en conflicto con ella (sección 4.4).

FIGURA 4.14

Cambio de actitud por reorganización de clases de equivalencia



Nota. Esquema original del autor, sin datos, en el formato de reorganización de clases formadas en laboratorio (como en Rosendo & de Melo, 2018). En los estudios de Carvalho y de Rose (2014) y Mizael et al. (2016a), el entrenamiento AB y BC crea una clase que entra en conflicto con la historia preexperimental.

La evidencia es mixta, y los resultados mejor sustentados son los de reorganización parcial. Rosendo y de Melo (2018) estudiaron clases con estímulos abstractos, adjetivos y profesiones asociadas a los géneros masculino y femenino. En la Etapa 1, dos grupos experimentales de ocho participantes cada uno fueron expuestos a entrenamientos y pruebas de formación de clases por igualación a la muestra demorada, y la formación de clases se verificó en los dos. En la Etapa 2, solo el Grupo Experimental 1 recibió entrenamientos de reversión, que redisponen las contingencias aprendidas en la Etapa 1, y pruebas de reorganización de clases. Un grupo de control, de 32 participantes, no pasó por ningún entrenamiento ni prueba. Los tres grupos evaluaron las profesiones y los símbolos abstractos en un diferencial semántico. En el Grupo Experimental 1, los participantes alcanzaron el criterio de reorganización, del 91% de aciertos, solo en las relaciones de simetría; en las demás pruebas, las puntuaciones quedaron cerca del 83% en seis participantes. En comparación con el control, las evaluaciones en el diferencial semántico fueron parcialmente coherentes con las clases y con los entrenamientos de reversión, con transfe-

rencia de función a algunas escalas, y las autoras concluyen que es necesario mejorar los controles sobre la historia preexperimental con los estímulos y sobre la cantidad de relaciones revertidas.

Con rostros y niños, la tabla de la sección 4.4 ya mostró la variación: uno de cuatro niños en de Carvalho y de Rose (2014), 13 de 13 con entrenamiento modificado en Mizael et al. (2016a), 33 de 46 en Mizael et al. (2021b). Dos puntos interesan a la lectura política. Uno es el mantenimiento: en Mizael et al. (2021b), 27 de los 46 niños mantuvieron al menos una relación después de seis semanas, y el grupo de control mantuvo menos. El otro es el tiempo: Mizael et al. (2016b) sostienen que un entrenamiento de una o dos horas es, probablemente, insuficiente para modificar relaciones preexperimentales ampliamente reforzadas por la comunidad verbal. Es una hipótesis de los revisores, no un dato de un estudio que la haya puesto a prueba.

Los cómics como procedimiento de cambio

Los estudios del Instituto Walden4 sustituyen el entrenamiento de clases, cuando el objetivo es cambiar una actitud, por un cómic (una historieta). La razón, para los autores de uno de los volúmenes (Moreira et al., 2022a), es la resistencia: los métodos tradicionales de cambio de actitud no toman en cuenta la resistencia actitudinal, que los métodos basados en narrativas sí consideran, y las narrativas, al tratar el prejuicio de forma más sutil, diluida en la trama, tienen la capacidad de minimizar la percepción de intención persuasiva y permiten que el lector se identifique con los personajes. Los autores no encontraron, en la literatura, experimentos sobre la formación de actitudes después de la lectura de cómics con el paradigma de equivalencia y el diferencial semántico, y es ese el vacío que el Estudio 1 del volumen se propone llenar.

Ese estudio, de Elisa de Pennafort Rabelo, tiene el diseño que repiten los siguientes. Siete estudiantes de cualquier carrera universitaria, excepto Psicología, de la que solo se admitían estudiantes del primer semestre, participaron en dos sesiones, una para cada historia, con un intervalo de tres días como máximo, por Google Meet y por Google Forms (Moreira et al., 2022a). Cada historia, hecha en Pixton, tiene siete viñetas y tres personajes mujeres: dos en conflicto de convivencia en la casa y una en un papel neutro; en la primera historia, Amanda es el único personaje negro. Antes y después de la lectura, el participante evaluó a los tres personajes de cada historia (neutro, víctima y agresor) en una escala de diferencial semántico de 13 pares de adjetivos, de 1 a 7. La lectura produjo efecto en las evaluaciones de los personajes de todos los participantes, y la mayoría atribuyó valores más positivos a las víctimas, más neutros a los personajes neutros y más negativos a los agresores, sobre todo

después de la segunda historia, lo que puede indicar interferencia de la primera sesión. Los autores no pudieron verificar la influencia de sesgos raciales y proponen, para eso, reproducir el experimento con nuevos participantes, cambiando el orden de presentación de las historias. La Figura 4.17 muestra, en la primera columna, el elenco de ese estudio, y la Figura 4.18, el conflicto de la primera historia.

El Estudio 2 del mismo volumen, de Ingrid Taveira Premuli, usa cómics en un tema que interesa a la lectura política: el autoritarismo. Las historias tenían personajes que representaban los tres factores de la Escala de Autoritarismo de Derecha (autoritarismo, cuestionamiento de la autoridad y tradicionalismo), y el estudio investigó, por igualación a la muestra, la transferencia de función de estímulos con significados de autoritarismo a estímulos abstractos. Según el resumen, el 100% de la muestra formó clases de equivalencia y el 80% tuvo las evaluaciones influidas por las historias; la ambivalencia en la evaluación, en algunos participantes, lleva a los autores a cuestionar la construcción y la validez de la escala de diferencial semántico (Moreira et al., 2022a).

El estudio político de la serie es el Experimento 8 de la Parte I (Moreira et al., 2022c), de Luiz Renato Valio Freire, y la Figura 4.15 lo muestra entero. Cinco adultos leyeron, en tres tramos, una historia de 18 viñetas con tres personajes: C1, con un ideal político orientado a la izquierda; C2, a la derecha; y C3, con opiniones neutras, que sirve para observar variables que estén interfiriendo en la investigación. Los personajes y las viñetas se hicieron en Storyboard Maker, y la historia usa pocos escenarios, con pocos elementos, para reducir influencias ajenas a lo que se estudia. En el primer tramo (Q1), las tres protagonistas se encuentran y conversan sobre el tiempo y la posibilidad de lluvia; es la ambientación, y la evaluación que sigue mide la primera impresión. En el segundo (Q2), C1 y C2 expresan opiniones opuestas sobre el armamento civil y van a marchas opuestas, una a favor y otra en contra. En el tercero (Q3), las tres comparten un auto de transporte por aplicación y aparecen en escena dos banderas de movimientos político-ideológicos, que, en la figura, se leen como la bandera de Brasil junto a C2 y la de la hoz y el martillo junto a C1. Después de cada tramo, el participante evalúa a C1, C2 y C3 en los mismos 13 pares de adjetivos, de 1 a 7. La comparación entre las tres pruebas es la medida del cambio, con el sujeto como su propio control. Cada participante tardó en promedio 20 minutos, en una recolección de una semana, y el experimento empezó solo después de la aprobación del Consejo de Ética.

El método tiene una ambigüedad, que conviene registrar. La primera etapa se llama pretest y se describe como la presentación de los rostros de los personajes seguida de la escala, antes de que el participante conozca la ideología polí-

tica de estos, pero la descripción de los tramos dice que Q1 se usó antes de la primera evaluación. La Figura 4.15 sigue esta segunda descripción, la más detallada; en las dos, la primera prueba viene antes de cualquier elemento político, que la descripción de los estímulos excluye de la primera evaluación.

Para los gráficos, los autores convirtieron las marcas de 1 a 7 en valores de −3 a +3 y reorganizaron los pares que se habían presentado invertidos, de modo que los adjetivos negativos quedaran en la parte inferior; la línea azul es la Prueba 1, la roja, la Prueba 2, y la amarilla, la Prueba 3. La Figura 4.16 reproduce el gráfico del Participante 1. En C2, las líneas de las Pruebas 2 y 3 casi se superponen: lo que cambió después de las opiniones políticas se mantuvo después de las banderas; en la mayor parte de los pares, las dos quedan por debajo de cero, y en agradable–desagradable llegan a −3. En C1 y C3, los cambios fueron pocos, con excepciones, y, en conjunto, el participante tendió a evaluar a C1 de forma más positiva y a C2 y C3 de forma más negativa a lo largo de las tres evaluaciones.

Entre los cinco participantes, hubo cambios de marca para todos los personajes. C1 recibió evaluaciones más positivas que las otras dos, y C3, el personaje neutro, recibió evaluaciones mayoritariamente negativas de los Participantes 2, 3, 4 y 5. Los Participantes 3 y 5 cambiaron de un polo de la escala al otro, y el 2 mantuvo evaluaciones más parecidas entre las pruebas. En conversaciones espontáneas después de la aplicación, que los autores presentan como evidencia anecdótica, los participantes dijeron que C3 fue evaluada de modo más negativo por su pasividad y por la falta de interacción con los otros personajes, y que el personaje con una ideología más parecida a la del participante era mejor evaluado. Para el análisis de la conducta, el primer relato es instructivo: el personaje planeado como control neutro no funcionó como control, porque, para quien evalúa, no actuar también es algo que se evalúa. El segundo sugiere que la semejanza entre quien evalúa y quien es evaluado puede contar tanto como lo que muestra la historia, y un estudio con cinco participantes y sin grupo de comparación no separa una cosa de la otra. Los autores ven el resultado en consonancia con los dos estudios anteriores del grupo, el de Rabelo y el de Premuli, en cuanto a que la lectura de cómics puede generar cambios de actitud hacia personajes ficticios, y recomiendan un método mixto, con entrevistas individuales sobre los motivos de cada evaluación.

FIGURA 4.15

El Experimento 8 de Moreira et al. (2022c): tres personajes, tres tramos de la historia, tres evaluaciones

1. El elenco: tres personajes



C1
ideal político orientado a la izquierda



C2
ideal político orientado a la derecha



C3
opiniones neutras (sirve de control)

2. Tres tramos de la historia, tres evaluaciones (13 pares de adjetivos, de 1 a 7)

1





Q1 · ambientación

Las tres amigas se encuentran y hablan del tiempo y de la posibilidad de lluvia. Todavía no hay opinión política.
Prueba 1: la primera impresión de cada una.

2





Q2 · opiniones políticas opuestas

C2 va a la marcha a favor del armamento civil; C1, a la marcha en contra. C3 no toma partido.
Prueba 2: ¿las ideas cambian la evaluación?

3





Q3 · dos banderas

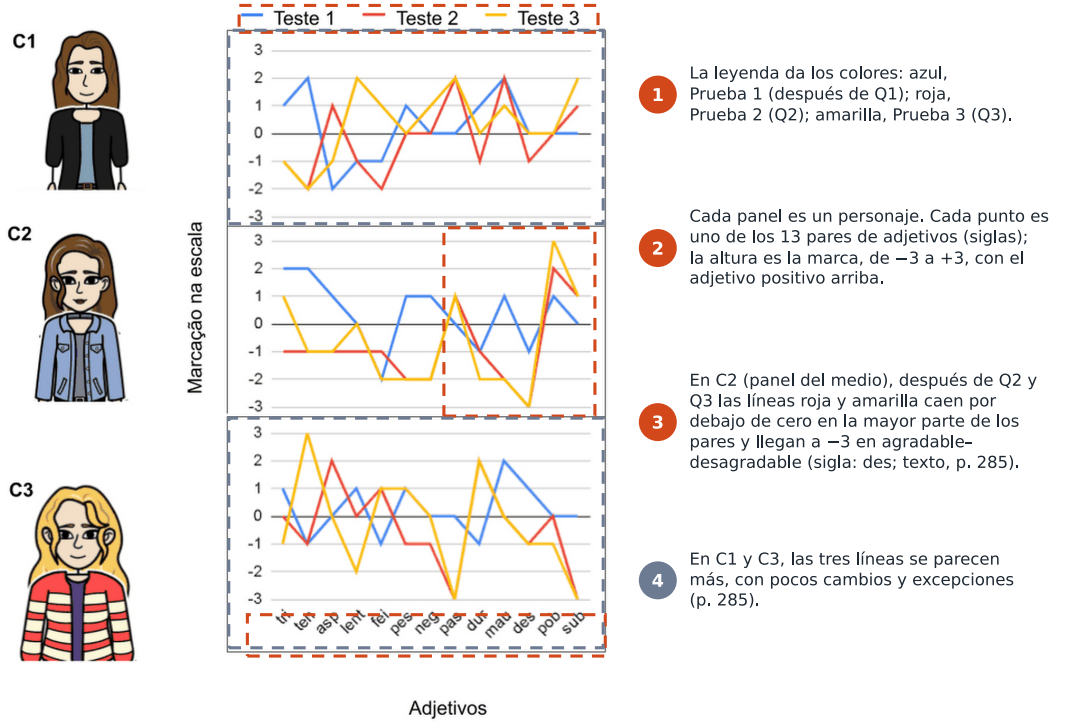
Las tres comparten un auto de transporte por aplicación; C2 aparece con la bandera de Brasil y C1 con la de la hoz y el martillo.
Prueba 3: ¿los símbolos cambian la evaluación?

En cada prueba, el participante evalúa a C1, C2 y C3 en los mismos 13 pares de adjetivos. La diferencia entre las pruebas es la medida del cambio; cada participante es su propio control.

Nota. Personajes y viñetas reproducidos de Moreira et al. (2022c); la organización en tres momentos y los textos laterales son esquema del autor, a partir del método. Los colores de las pruebas son los de las líneas de los gráficos del estudio (Figura 4.16). La atribución de las banderas a los personajes se leyó en la última imagen; el texto del estudio dice solo que son dos banderas. Los globos de diálogo están en portugués, como en el estudio original; los textos laterales resumen cada tramo.

FIGURA 4.16

Las evaluaciones del Participante 1 en el Experimento 8



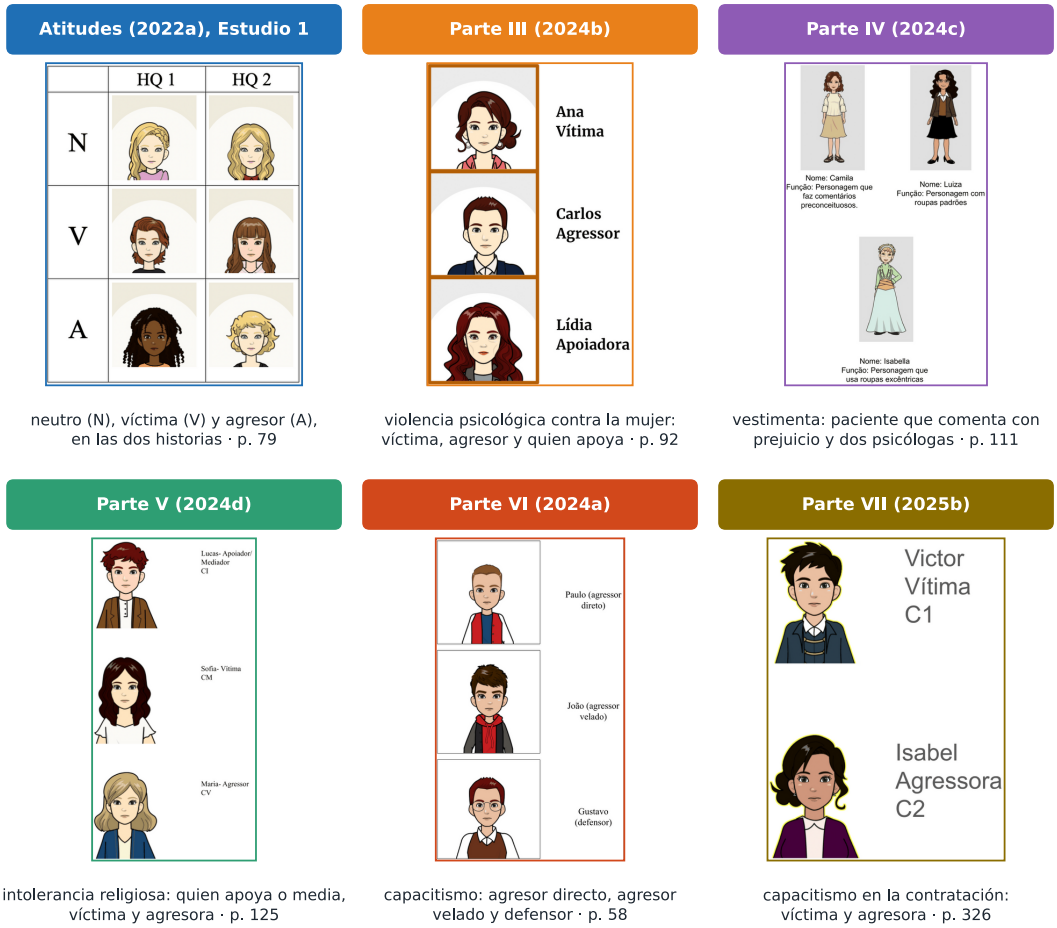
Reproducción de la Figura 4 de Moreira et al. (2022c, p. 286).

Nota. Reproducción de la Figura 4 de Moreira et al. (2022c), con los destacados y la guía añadidos. Las siglas del eje son las del estudio (tri, ten, asp, lent, fei, pes, neg, pas, dur, mau, des, pob, sub), en el orden de los 13 pares de adjetivos, con el polo negativo nombrado; el estudio no las desarrolla, y el texto lee “des” como agradable–desagradable. Los rótulos están en portugués, como en el estudio original: las siglas abrevian los adjetivos en portugués (“des”, de desagradável), y la leyenda y los títulos de los ejes dicen Teste 1–3 (Pruebas 1–3), Marcação na escala (marca en la escala) y Adjetivos (igual que en español).

Los volúmenes siguientes repiten el diseño con otros temas y otros elencos. En las Partes IV, V y VI, cada estudio trae, como el Experimento 8, personajes, trama, guion, el formulario de diferencial semántico y pruebas antes y después de la lectura (Moreira et al., 2024c, 2024d, 2024a, índices). La Tabla 4.2 enumera los temas, y la Figura 4.17 muestra el elenco de seis estudios: por los rótulos impresos en las figuras, suele incluir a quien sufre el acto, a quien lo practica y a quien apoya, defiende, media u observa. La Figura 4.18 muestra lo que lee el participante: tres escenas, de tres libros.

FIGURA 4.17

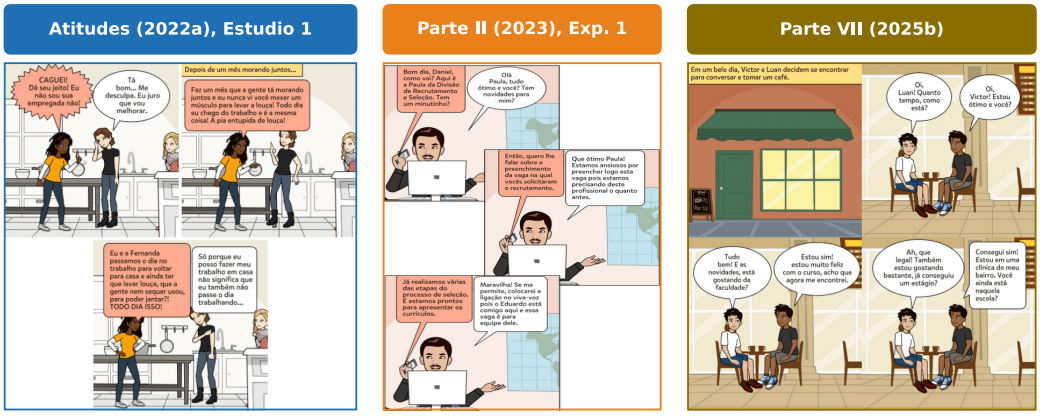
El elenco de seis estudios con cómics



Nota. Personajes reproducidos, con los rótulos de cada libro, de Moreira et al. (2022a; 2024b; 2024c; 2024d; 2024a; 2025b); las franjas y las leyendas son del autor. Los temas son los de los títulos de los estudios (Tabla 4.2). Los rótulos están en portugués, como en cada libro: Vítima (víctima), Agressor y Agressora (agresor, agresora), Apoiador y Apoiadora (quien apoya), Mediador (mediador), defensor (defensor) y velado (encubierto).

FIGURA 4.18

Tres escenas de cómics



conflicto entre compañeras de apartamento, viñetas 3 a 5 · p. 73

edadismo en el mercado de trabajo: escena 1, viñetas 1 a 3 · p. 102

capacitismo en la contratación: escena 1, viñetas 1 a 4 · p. 327

Nota. Reproducidas de Moreira et al. (2022a), del Experimento 1 de Moreira et al. (2023) y de Moreira (2025b), en ese orden. Lo que dicen los personajes importa aquí menos que la forma: pocos escenarios, parlamentos cortos y un conflicto que el participante lee antes de evaluar a los personajes. Los globos de diálogo y las leyendas están en portugués, como en los libros originales.

TABLA 4.2

Libros del Instituto Walden4 con estudios de formación y cambio de actitudes y los temas de los estudios

Libro	Temas de los estudios
Atitudes: histórias em quadrinhos, análises semânticas em mídias sociais e o paradigma de relações de equivalência de estímulos (Moreira et al., 2022a)	Lectura de cómics y personajes ficticios (Estudio 1); autoritarismo expresado por personajes (Estudio 2); trabajo doméstico (Estudio 3); conducta verbal en Twitter, movimiento Black Lives Matter (Estudio 4, Capítulo 6)
Preconceito: doze experimentos e um paradigma (Moreira et al., 2022b)	Control de estímulos y medición de actitudes (Experimento 1); transferencia de función y equivalencia con estímulos de significado social (Experimentos 2 a 12, sección 4.4)
Estudos sobre histórias em quadrinhos e atitudes, Parte I (Moreira et al., 2022c)	Autismo en la escuela y en el mercado de trabajo; transfobia; postura científica o pseudocientífica; psicólogos con credo religioso; atractivo físico; exreclusos; prejuicio político (Experimento 8); xenofobia; mujeres en los videojuegos; cirugía plástica e influenciadoras digitales; ostentación de riqueza por celebridades
Parte II (Moreira et al., 2023)	Validez y precisión del instrumento de diferencial semántico; edadismo en el mercado de trabajo; comunicación agresiva, asertiva y pasiva; obesidad infantil; vestimenta; TEA en el mercado de trabajo; prejuicio e intolerancia religiosa; discriminación racial

Libro	Temas de los estudios
Parte III (Moreira et al., 2024b)	Violencia psicológica contra la mujer en el ambiente universitario; pseudociencias; apelación emocional; géneros musicales; transfobia
Parte IV (Moreira et al., 2024c)	Transfobia; homofobia; discriminación por la vestimenta; género en el deporte de alto rendimiento; gordofobia
Parte V (Moreira et al., 2024d)	Xenofobia; conducta prejuiciosa e intolerancia religiosas
Parte VI (Moreira et al., 2024a)	Esquizofrenia en el mercado de trabajo; capacitismo contra personas autistas; inclusión de niños con TEA; ambivalencia en una narrativa en cómic
Parte VII (Moreira, 2025b)	Homofobia y ambivalencia; racismo y homofobia en el fútbol; noticias de chismes; personas en situación de calle; madres en entrevistas de trabajo; tratamientos alternativos para el autismo; estímulos emocionales; contratación de personas con TEA; redes sociales; natación inclusiva; el caso Gabriela Pugliesi; exadictos

Nota. Temas según los índices de los volúmenes; "Parte II" a "Parte VII" abrevian el título Histórias em quadrinhos e o estudo das atitudes. De los estudios de la tabla, solo dos se leyeron en detalle en este capítulo, el Estudio 1 de Moreira et al. (2022a) y el Experimento 8 de Moreira et al. (2022c).

El volumen organizado por Moreira (2025a) reúne estudios de sujeto único, mayoritariamente con universitarios, con igualación a la muestra y con cómics, sobre sexismo, homofobia, edadismo, transfobia, gordofobia y prejuicio contra personas con síndrome de Down y con autismo, en los que la actitud se mide, sobre todo, con una escala de diferencial semántico. El resumen del volumen relata que la igualación formó clases y transfirió funciones evaluativas, en especial a estímulos inicialmente neutros o poco cargados emocionalmente, y que hubo resistencia al cambio de prejuicios profundamente arraigados o asociados a una fuerte carga emocional (resumen del volumen).

DATO

En estudios de reorganización con estímulos de significado social, la reorganización fue parcial (Rosendo & de Melo, 2018, solo para la simetría), varió con los parámetros del entrenamiento (Mizael et al., 2016a, 2021b) y se mantuvo en parte después de seis semanas (27 de 46). En los dos estudios con cómics leídos en detalle, hubo alteraciones en las marcas de todos los participantes (Moreira et al., 2022a, Estudio 1: siete; 2022c, Experimento 8: cinco), y el personaje planeado como neutro en el Experimento 8 recibió evaluaciones mayoritariamente negativas de cuatro de los cinco.

INTERPRETACIÓN	Reorganizar una clase preexperimental es posible en condiciones de entrenamiento optimizadas, pero el efecto es parcial, medido en una escala justo después del entrenamiento, y no dice lo que la persona dirá en otro contexto. Los estudios con cómics examinan el efecto de la información sobre la evaluación de personajes ficticios; con el sujeto como su propio control (Moreira et al., 2022c), el cambio en las marcas no se distingue, por sí solo, de un efecto de repetición de la medida, y el Estudio 1 de Moreira et al. (2022a) registra una posible interferencia de la primera sesión sobre la segunda.
HIPÓTESIS	El cambio de actitud en el discurso político dependería, además del entrenamiento, de contingencias que sigan reforzando las relaciones nuevas y dejen de reforzar las antiguas; las relaciones antiguas siguen siendo reforzadas por la comunidad verbal fuera del laboratorio. Es lo que el Capítulo 12 trata como intervención, no como un dato de este.

4.8 Medidas, interpretaciones rivales y límites

La medida de actitud más frecuente en los estudios de este capítulo es una escala: de diferencial semántico, con pares de adjetivos bipolares, de siete puntos en el artículo con presidentes y en barra deslizante de 0 a 100 en el libro, y de siete puntos en los estudios con cómics y con género y profesiones; Likert, en Dixon y Lemke (2007) y en Cameoka y Moreira (2021). Los autores del estudio con presidentes la toman como medida de actitud porque relaciona adjetivos con estímulos (Pereira & Moreira, 2021b). Hay medidas de otro tipo: la prueba de asociación implícita, usada por Weinstein et al. (2008), en la que el autoinforme no se correlacionó con la respuesta, y medidas basadas en la latencia, que Mizael et al. (2021a) incluyen entre las contribuciones brasileñas a la evaluación de sesgos raciales. Ninguna de ellas deja de ser conducta bajo control del procedimiento que la produce.

Desde el punto de vista analítico-conductual, una marca en la escala es conducta verbal dirigida a un investigador. Guerin (1994) observa que las escalas y las preguntas dependen de quién pregunta y de la cultura verbal en la que se hace la investigación, porque las actitudes dependen siempre de consecuencias sociales, y que la audiencia funciona como estímulo discriminativo para informar actitudes específicas, de modo que una pregunta puede “mandar” la respuesta. De ahí la expectativa de inconsistencias entre lo que se declara y lo que se hace, sobre todo en la conducta verbal. La consecuencia es directa para

los experimentos anteriores: un cambio en la marca es un cambio en lo que la persona marca, ante ese investigador y esa escala, y no un cambio demostrado en lo que hará o dirá ante otras audiencias.

El instrumento de los dos trabajos con presidentes tiene dos peculiaridades que piden cautela. La primera es que los números de uno y otro no se comparan directamente. En el artículo, la escala tenía siete casillas por par, marcadas en papel y convertidas en números de −3 a +3; en el libro, era una barra deslizante de 0 a 100, respondida en la computadora, con pares de polos alternados, y la semejanza entre estímulos se calculó de otro modo, como 100 menos la diferencia entre las notas. El GE y el GT del artículo y la semejanza del libro miden, a grandes rasgos, cuánto se parecen dos evaluaciones, pero las cuentas son diferentes. La segunda es una inconsistencia del artículo: el método dice que, en el par Derecha–Izquierda, Derecha quedó con −3 e Izquierda con +3, y la discusión describe las fotografías de Bolsonaro y de Lula como, respectivamente, derecha (valores próximos a 3) e izquierda (valores próximos a −3). Los signos de la discusión son los inversos de los del método. Este libro registra la divergencia y no la resuelve, porque no tiene acceso a los datos brutos, y el punto que usa el capítulo, el de que los participantes evaluaron sistemáticamente las fotografías como derecha, izquierda y centro, no depende del signo.

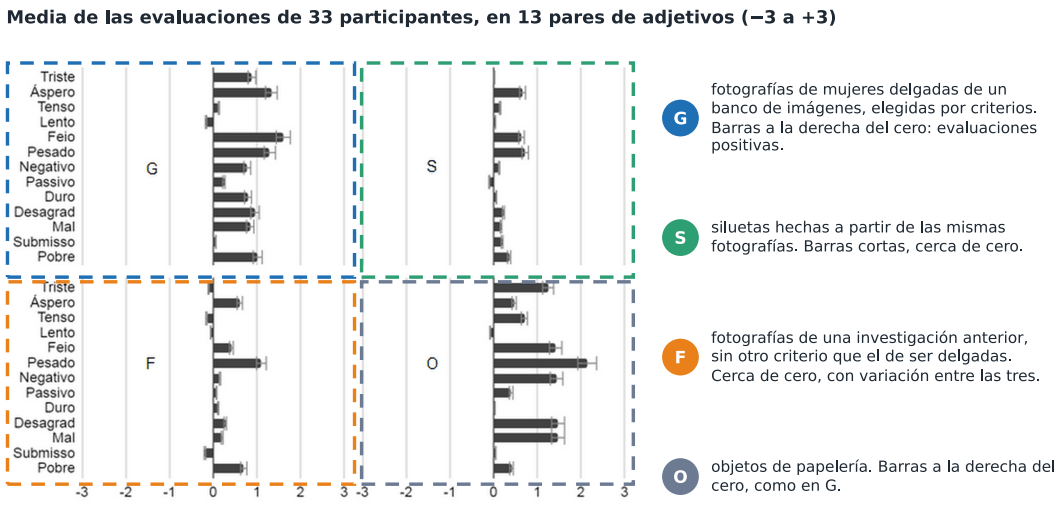
La selección de estímulos también controla la medida. Cameoka y Moreira (2021) hicieron tres estudios, con 75, 40 y 30 universitarios, en los que evaluaron, en una escala de 1 (malo) a 10 (bueno), fotografías de hombres blancos y de hombres negros. La diferencia entre las medias de los dos grupos de rostros fue significativa en los Estudios 1 y 2 y no en el Estudio 3, en el que se usaron los rostros de blancos mejor evaluados y los de negros peor evaluados en los estudios anteriores; y la media de los rostros de blancos y de negros aumentó, de modo significativo, del Estudio 1 al 2, en el que los rostros estaban discretamente sonrientes. Los autores suponen demostrar, con esto, el potencial de distorsión de una selección de estímulos poco rigurosa. La consecuencia es que "la actitud hacia X" medida con una fotografía de X es, en parte, la actitud hacia esa fotografía.

El Experimento 1A de Moreira et al. (2022b) muestra el problema en un diseño hecho para ello. Treinta y tres personas, de 15 a 53 años, evaluaron en un formulario en línea 12 estímulos en 13 pares de adjetivos, de 1 a 7: tres fotografías de mujeres delgadas de un banco de imágenes, elegidas por criterios (en lencería o bikini, de frente, edad aparente de 20 a 30 años, fondo blanco y carácter comercial, y semejanza, juzgada por ocho jueces, con las siluetas 2 o 3 de la Escala de Stunkard); tres fotografías de mujeres delgadas de una investigación anterior, sin otro criterio; tres siluetas, hechas a partir de las fotogra-

fías del primer conjunto, con el cuerpo pintado de negro; y tres objetos de papelería. Cada estímulo apareció dos veces, en orden semialeatorio. Las fotografías elegidas por criterios recibieron evaluaciones positivas, iguales o mayores que 0,8 en la mayor parte de los pares; las siluetas, evaluaciones cerca de cero, en general entre 0 y 0,4; y la diferencia fue estadísticamente significativa en todos los pares, excepto Tenso/Relajado y Sumiso/Dominante (Figura 4.19). Entre las tres fotografías sin criterio, una recibió evaluaciones más positivas, otra más neutras y la tercera más negativas, y los autores atribuyen la variación al único criterio usado, el de ser delgadas, a juicio del investigador. Como las siluetas conservan la forma del cuerpo y quedaron cerca de cero, los autores sugieren que las respuestas a las fotografías no estaban bajo control de la forma del cuerpo, sino de otras partes de los estímulos, y concluyen que la estandarización en la selección de los estímulos es esencial para que las respuestas queden bajo control de la variable programada, y que el color de la piel o del cabello, cuando no es la variable de interés, puede crear sesgos evaluativos. Para investigaciones futuras, proponen dos estudios separados, uno de selección de los estímulos y otro para la tarea de discriminación o de reversión de equivalencia. Para la lectura de declaraciones políticas, la consecuencia es la misma: el ejemplar elegido para representar a un grupo es lo que se evalúa, y la actitud "hacia el grupo" es una inferencia hecha a partir de él.

FIGURA 4.19

El mismo concepto, cuatro tipos de estímulo, cuatro perfiles de evaluación



Reproducción de la Figura 3 de Moreira et al. (2022b, p. 47).

Cada barra es la media de un par de adjetivos; a la derecha del cero está el adjetivo positivo. Lo que cambia es el estímulo, no el concepto: G, S y F pretendían representar mujeres delgadas (pp. 41, 51).

Nota. Reproducción de la Figura 3 de Moreira et al. (2022b), con los destacados y la guía añadidos; cada barra es la media de los participantes para un par de adjetivos, y los trazos finos son los de la figura original (el texto no dice qué medida representan). Los rótulos de los pares son los del gráfico, en portugués.

Queda lo que ninguna de estas medidas alcanza. Los experimentos tratan de conducta verbal individual ante estímulos elegidos por el investigador, en sesiones cortas y con muestras pequeñas, de 4 a 50 participantes en los estudios de la Tabla 4.1 y de cinco por experimento en los de presidentes, muchas veces estudiantes universitarios o niños reclutados en escuelas. Passarelli et al. (2023) distinguen el racismo en tres niveles, estructural, institucional e individual, y señalan la escasez de trabajos experimentales, en parte por la dificultad de definirlo en términos conductuales; lo que este capítulo examinó es el nivel individual. El paso del laboratorio al discurso público es por analogía, y aquí se asume como tal: el paradigma ofrece un modelo de cómo las funciones evaluativas llegan a nombres sin contacto directo y de cómo la historia social compite con lo que se enseña, y con ese modelo leen los Capítulos 5 y 7 las declaraciones.

Interpretaciones rivales para el mismo dato de transferencia

Emparejamiento directo. En el diseño de Pereira y Moreira (2021b), las figuras B y C se relacionaron directamente con la fotografía (entrenamientos AB y AC); la relación derivada es BC. El condicionamiento evaluativo, en el que el emparejamiento directo entre un estímulo neutro y otro con valencia cambia la respuesta evaluativa al neutro (Mizael et al., 2016b), explica la semejanza entre A y B, y entre A y C, sin recurrir a la clase. Los autores proponen poner a prueba series lineales frente a uno-a-muchos, con la distancia nodal; Bortoloti y de Rose (2009) ya mostraron que las figuras a un nodo de los rostros fueron juzgadas semejantes y a tres nodos, no.

Control por la audiencia. La marca puede reflejar lo que el participante supone que el investigador espera (Guerin, 1994).

Línea de base no neutra. Las evaluaciones neutras no son ausencia de significado (Pereira & Moreira, 2021b), y P02 evaluó las figuras abstractas, en el pretest, de modo positivo y negativo; la diferencia entre pretest y postest mezcla, entonces, la transferencia con lo que las figuras ya eran para él.

Variación de medida repetida. Los autores admiten no saber cuánto de la diferencia en estímulos con significado previo es variación entre aplicaciones de la escala.

Síntesis

En este libro, la actitud es una evaluación de un objeto que es, al mismo tiempo, conducta verbal: mantenida por consecuencias dispuestas por la comunidad verbal y capaz de funcionar como tacto, intraverbal o mando. El paradigma de equivalencia de estímulos ofrece el modelo de cómo una función evaluativa llega a un estímulo sin contacto directo: después de que una clase se

forma por discriminaciones condicionales, la función de un miembro aparece en los demás, en pruebas sin consecuencias diferenciales, y varía con la distancia nodal.

La literatura con estímulos de significado social muestra un patrón. Con estímulos arbitrarios, las funciones evaluativas aparecen con un tamaño pequeño después de entrenamientos cortos. Con estímulos que contrarían la historia del participante, la formación de la clase se interrumpe en parte de los casos, y la interrupción acompaña la fuerza de la historia, que los estudios comparan (entre edades, entre sexos), pero no manipulan; con entrenamientos modificados, niños formaron la clase pretendida y la mantuvieron en parte. Los experimentos con presidentes de Brasil muestran el mecanismo en estudiantes universitarios, en un diseño de pretest y postest sin control, con cinco participantes por experimento. En los cinco del artículo, las evaluaciones de las figuras abstractas pasaron a parecerse a las de las fotografías (GE superior al 90%), con cambios moderados (GT del 12% al 49% en las figuras B); en los diez del libro, la semejanza aumentó en promedio de 5 a 14 puntos porcentuales, más para figuras abstractas que para rostros neutros. Los dos trabajos sugieren que el rótulo derecha–izquierda ya forma parte de la clase con la que cada nombre llega al participante. En los estudios del Instituto Walden4 con cómics, las marcas cambiaron en todos los participantes (siete en un estudio, cinco en otro), con el sujeto como su propio control y sin grupo de comparación; en el estudio político, el personaje planeado como neutro no funcionó como control, y un experimento sobre la selección de estímulos mostró que el tipo de fotografía, y no solo el concepto, controla la evaluación.

Todos estos datos son marcas en una escala o elecciones en una pantalla, dirigidas a un investigador, y admiten al menos dos lecturas rivales: el emparejamiento directo entre fotografía y figura, sin clase, y el control por la audiencia. Las relaciones de oposición entre “nosotros” y “ellos” son una hipótesis prometedora para leer el discurso político, apoyada por pocas entrevistas y sin datos sobre declaraciones. El Capítulo 5 usa lo que este capítulo estableció sobre la transferencia para el rótulo de grupo, y el Capítulo 7, para los nombres de candidatos.

Referencias

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. (Ed.). (2025a). *Formação e mudança de atitudes: Pareamento ao modelo e história em quadrinhos* (1.ª ed.). Instituto Walden4.

Moreira, M. B. (Ed.). (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* (1.ª 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., Morais, 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., Povoa, 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* (1.ª 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* (1.ª 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* (1.ª 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* (1.ª 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

instituto WALDEN4 | ACADEMIA WALDEN4

PRÁCTICAS CON EVIDENCIA

Lo que funciona en el autismo, práctica por práctica

Una página para cada práctica con evidencia, a partir de la revisión del National Clearinghouse on Autism Evidence and Practice: 31.779 artículos cribados y 972 analizados, con el número de estudios que sustenta cada una.

28 prácticas descritas

Acceda, gratis y sin registro

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

CAPÍTULO

5

Discurso de odio: mando disfrazado y control aversivo

En 2011, el programa CQC, de la cadena TV Bandeirantes, le preguntó al entonces diputado Jair Bolsonaro qué haría si tuviera un hijo gay. Según la Agência Lupa (agencia brasileña de verificación de datos), que reproduce en su informe de 2026 la verificación publicada el 3 de agosto de 2018, él respondió: "Eso ni me pasa por la cabeza, porque ellos tuvieron una buena educación. Soy un padre presente, así que no corro ese riesgo" (Agência Lupa, 2026, p. 24). En noviembre de 2017, el Tribunal de Justicia de Río de Janeiro confirmó la condena del diputado por daño moral colectivo, por declaraciones de contenido homofóbico hechas al programa; la pena fijada en 2015 preveía una indemnización de R$ 150 mil, y el proceso, en la fecha de la verificación, aún estaba en trámite. "Contenido homofóbico" es la descripción de Lupa, y la condena es del tribunal, según Lupa. Ambas se registran aquí con esa atribución, sin clasificación añadida.

Este capítulo parte de casos como ese, el de un enunciado que una instancia externa clasificó, y pregunta qué puede decir el Análisis de la Conducta sobre él sin repetir la clasificación como si fuera una explicación. Llamar a un enunciado discurso de odio, homofóbico o racista describe lo que es para quien lo clasifica. No dice qué lo produce ni qué lo mantiene. La pregunta del capítulo es la segunda: qué contingencias mantienen enunciados de este tipo, del lado del hablante, de la audiencia y del grupo al que se refieren. Los instrumentos son tres: las contingencias entrelazadas (sección 5.1), el mando disfrazado (5.2) y el control aversivo (5.3). Después vienen el rótulo del grupo y la transferencia de función (5.4), lo que los datos permiten afirmar sobre el paso del discurso al acto (5.5) y la ficha del Capítulo 3 aplicada al caso de la apertura (5.6).

Una regla de todo el libro vale aquí con más fuerza que en los otros capítulos: "discurso de odio" es un rótulo externo. Solo se llama así al enunciado clasificado por una instancia externa, como una decisión judicial, la remoción por una plataforma o la codificación de un estudio científico, con la instancia nombrada. Sin clasificación externa, se describe la función del enunciado y no se lo rotula. Las definiciones y las medidas que usa cada estudio se tratan en el Capítulo 1; aquí se toman tal como el estudio las presenta.

5.1 Lo que produce la conducta del hablante: aprobación, retirada de amenazas y funciones que varían con el contexto

Para saber qué mantiene un enunciado, el análisis pregunta por las consecuencias que produce. Lo que las fuentes verificadas permiten es una lista corta de candidatas, ninguna puesta a prueba con declaraciones de políticos.

La primera es la aprobación o la atención de una audiencia. Es el reforzamiento social generalizado del Capítulo 3 (sección 3.3): para Guerin (1994), las actitudes son enunciados mantenidos por ese tipo de reforzamiento en una comunidad verbal. Fazzano y Gallo (2015), que proponen una hipótesis funcional para la agresión contra homosexuales (sección 5.3), plantean como hipótesis, además del reforzamiento negativo, la existencia de reforzamiento social, proporcionado por los pares, por medio de la aprobación del grupo. La clase de respuestas que analizan tiene topografías diversas (golpear, insultar, torturar, apalear, excluir, discriminar) y se define por la función, y no por la forma.

La segunda es la remoción o la reducción de estimulación aversiva. Skinner (1957) incluye el escape y la evitación de estimulación aversiva entre los reforzadores generalizados, y entre los modos de estimular a alguien aversivamente enumera muchas clases de "daño verbal". Fazzano y Gallo (2015) plantean como hipótesis que la presencia de homosexuales sería aversiva para el agresor y que la agresión sería reforzada negativamente, y observan que la reducción de la intensidad del estímulo ya basta para el reforzamiento, de modo que la violencia no necesita terminar en muerte para ser reforzada. Que la aversividad sea aprendida y no propia del estímulo es su punto: no hay, en el propio estímulo, algo que sea aversivo, y el caso parece ser de aprendizaje. Los procesos de adquisición que señalan son el moldeamiento por reforzamiento diferencial, el seguimiento de instrucciones y la imitación de un modelo, "muy posiblemente" los tres juntos.

La tercera es que las funciones varían. Mizael (2018), en una revisión de la producción analítico-conductual sobre la homosexualidad (sección 5.3), al enumerar lo que el Análisis de la Conducta podría hacer por el tema, describe (citando un análisis contextual del racismo publicado por Guerin en 2005)

funciones diferentes de "ser racista" que controlarían la conducta de las personas: ser aceptado, aumentar el estatus del propio grupo, evitación, uso irreflexivo de estereotipos. Lo que la lista enseña es que el mismo enunciado puede tener, en ocasiones diferentes o en hablantes diferentes, consecuencias diferentes, y que ninguna de ellas exime de la pregunta por la ocasión.

Fazzano y Gallo (2015) critican las explicaciones que tratan la violencia como defensa contra los propios impulsos del agresor, o como defensa del yo, porque atribuyen la causa a una entidad "inaccesible e inmutable", no toman en cuenta los contextos por los que pasó el individuo y, en sus palabras, eximen de actuar a las agencias sociales. Mizael (2018) retoma el punto al proponer reducir los esencialismos que tratan la homofobia como una entidad dentro del individuo. El análisis funcional cambia la pregunta: de lo que hay dentro de quien dice a las condiciones en que el decir ocurre (Skinner, 1953/2003).

CONCEPTO EN USO

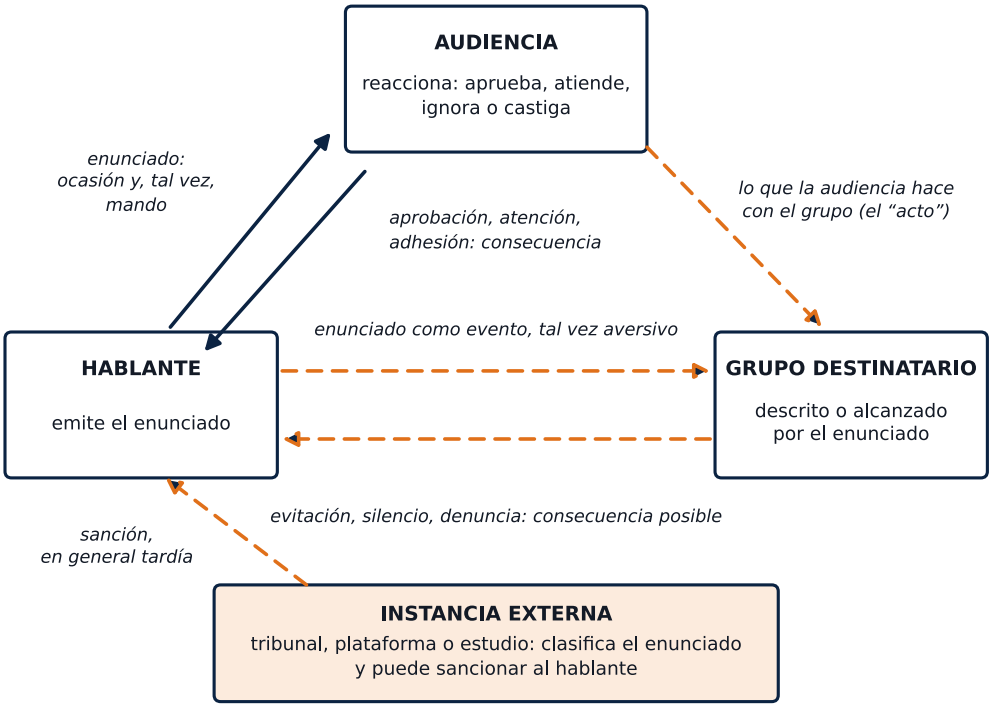
Contingencias conductuales entrelazadas (Glenn, 2004)

Las contingencias operantes están entrelazadas cuando la conducta de dos o más personas funciona como evento ambiental para la conducta de las otras (Glenn, 2004). El resultado de la conducta entrelazada no es el efecto acumulado de cada persona comportándose sola, sino el efecto de su conducta interrelacionada, "más que" o "distinto de" lo que cada una produciría aislada. En este capítulo, el entrelazamiento tiene tres participantes: el hablante, cuya respuesta es el enunciado; la audiencia, cuya reacción es consecuencia para la conducta del hablante; y el grupo destinatario, para el cual el enunciado es un evento y cuya reacción puede ser también consecuencia para la conducta del hablante. La aplicación a los tres se hace aquí, y no la hace Glenn.

La Figura 5.1 reúne las relaciones. Las dos líneas continuas, entre hablante y audiencia, son las que tienen algún dato indirecto, tratado en la sección 5.5: la declaración de líderes que precede a cambios en lo que hacen sus partidarios y la expresión pública que acompaña a la popularidad percibida de lo que se dice. Las discontinuas son las que el capítulo propone y ninguna fuente verificada mide: lo que el enunciado le hace al grupo destinatario, lo que el grupo destinatario hace de vuelta y lo que la audiencia hace con el grupo. La instancia externa no forma parte de la contingencia en el momento de la emisión: clasifica después y, cuando sanciona, lo hace con retraso. Solo pertenecerá a la contingencia en caso de recurrencia y de efecto supresor sobre respuestas de la misma clase que la sucedieron. En el caso de la apertura, la consecuencia documentada para el hablante vino de un tribunal, años después, y no de la audiencia del programa.

FIGURA 5.1

Hablante, audiencia y grupo destinatario: contingencias entrelazadas en torno a un enunciado



Línea continua: relación con dato indirecto en fuente verificada (sección 5.5). Discontinua: relación conceptual o hipótesis.

Nota. Esquema original del autor, a partir de Glenn (2004), Skinner (1957) y Fazzano y Gallo (2015). No representa datos.

5.2 Mando disfrazado: la opinión que funciona como instrucción

El mando, en el vocabulario del Capítulo 3, es el operante verbal cuya respuesta es reforzada por una consecuencia característica y que está, por eso, bajo control de privación o de estimulación aversiva (Skinner, 1957). Skinner distingue especies de mando por las consecuencias proporcionadas a la conducta del oyente al atender el mando del hablante. En la petición, el oyente ya está dispuesto a dar lo que se pide, y la respuesta solo indica lo que el hablante aceptará. En la orden, la respuesta del hablante establece una situación aversiva de la que el oyente solo escapa proporcionando la mediación adecuada: "¡Manos arriba!" no solo especifica una acción, sino que constituye una amenaza de la que la víctima escapa levantando las manos, y "¡La bolsa o la vida!" trae las dos partes, la que especifica el reforzamiento y la que nombra la consecuencia aversiva. Skinner escribe que las órdenes militares se obedecen por

una especie de amenaza permanente. Cuando el oyente gana algo propio, el mando es consejo, si hay reforzamiento positivo, o advertencia, si hay escape de estimulación aversiva; el permiso cancela una amenaza que lo contenía.

CONCEPTO EN USO

Mando disfrazado (Skinner, 1957; Guerin, 1994)

Skinner usa la expresión en dos pasajes. En el primero, a propósito de las audiencias negativas: la conducta verbal sobre el oyente, o sobre personas importantes para él, puede recibirse aversivamente, y con algunos oyentes pasamos a evitar los mandos o a usar mandos disfrazados en su lugar. En el segundo, el pregón "Call for Philip Morris" funciona como mando, aunque disfrazado de respuesta de un mensajero que llama a alguien. Skinner no desarrolla el concepto más allá de eso. Guerin (1994) lo acerca a lo que este capítulo necesita: los enunciados de actitud pueden traer consecuencias específicas y, así, calificar como mandos; el comentario de la jefa de que no le gusta quien toma café mientras trabaja no es un tacto inocente de una actitud privada y funciona como si ella hubiera mandado "¡No tome café mientras trabaja, o ya verá!", lo haya pretendido o no. Medeiros (2013) define los mandos disfrazados como respuestas verbales con topografía de tacto/intraverbal cuya ocurrencia está controlada por los reforzadores que mantienen los mandos, aunque su topografía no explicite claramente el reforzador. En este libro, *mando disfrazado* es el enunciado con forma de tacto o de opinión cuya función, según la evidencia disponible, es la de mando. La aplicación al discurso político es una hipótesis aquí propuesta.

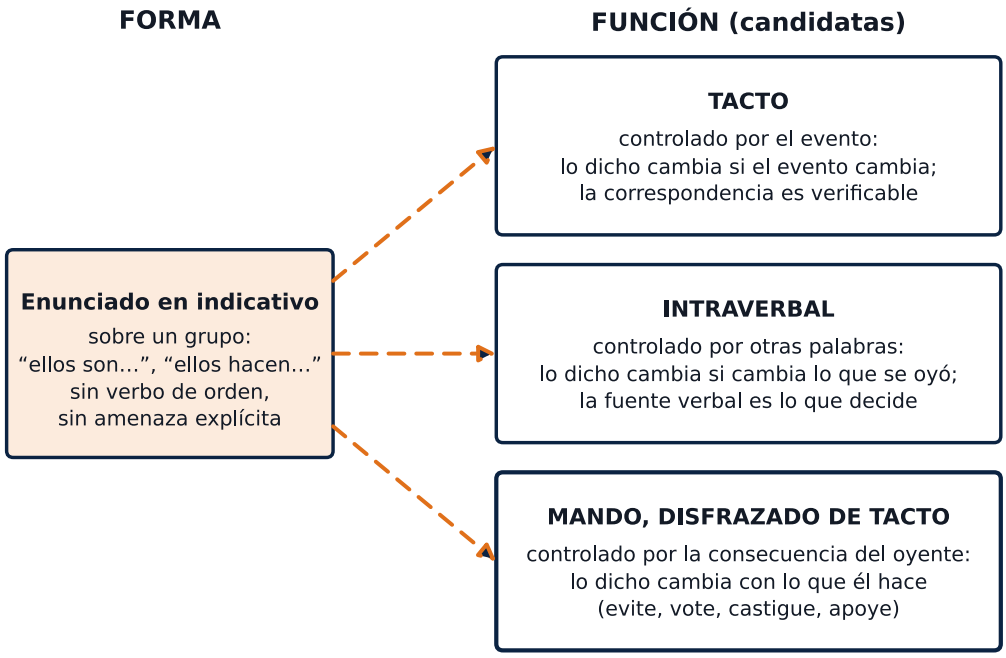
Los dos criterios que da el texto de Skinner son, por lo tanto, una condición de aparición y un criterio de identificación. La condición es una audiencia que castiga el mando explícito: con ella, el hablante pasa a usar la forma que no se castiga. El criterio es la divergencia entre forma y función: la forma es la de otra cosa, la función es la de mando. Este capítulo toma solo esos dos y la lectura de Guerin; el resto es aplicación.

La aplicación es esta. Un enunciado como "ellos son peligrosos" tiene la forma de un tacto, en indicativo, sobre un grupo. Si su función es la de mando a la audiencia, estará controlado por lo que el oyente hace después, y lo que se pide es: evítelos, vote en contra, apoye a quien los ataca, al menos no castigue. Nada en la forma dice eso. La Figura 5.2 muestra las tres funciones entre las que puede estar la misma forma, y que la sección 3.3 ya distinguió: tacto, intraverbal y mando, es decir, los tres operantes verbales cuya topografía no está determinada en una relación punto por punto con el estímulo antecedente verbal.

Lo que separa la hipótesis de mando de sus rivales son las tres preguntas que la ficha del Capítulo 3 ya hacía, aplicadas ahora a la conducta del oyente. Qué hace el oyente después: una respuesta observable y no una intención que se le atribuya. Qué hace el hablante cuando el oyente no atiende: Skinner observa que la repetición puede ser solo evidencia de fuerza excesiva de la respuesta y que, cuando la conducta verbal es ineficaz para alterar las circunstancias responsables de su fuerza, una respuesta puede ceder el paso a otra, de modo que, si la repetición pura se castiga, la forma varía (sección 3.5). Y si la forma cambia cuando cambia la audiencia: con una audiencia que castiga el mando explícito, el disfraz es esperable; con una audiencia que lo refuerza, lo que se esperaría es el mando explícito. En cuanto al autoclítico asertivo, que "insta al oyente a aceptar un estado de cosas dado" y que la sección 3.6 trató, Skinner ya lo clasifica como un tipo especial de mando (p. 326); "nunca", "jamás" y "siempre" llevan, por lo tanto, un mando incorporado que no depende del asunto.

FIGURA 5.2

Mando disfrazado: la misma forma de opinión, tres funciones candidatas



La forma es la misma; lo que la distingue es lo que la controla (Skinner, 1957, p. 36).

Nota. Esquema original del autor, a partir de Skinner (1957) y Guerin (1994). No representa datos. Las tres flechas son discontinuas porque las tres funciones son hipótesis para cualquier enunciado particular.

Ninguna de las fuentes verificadas registra las tres cosas para un enunciado clasificado como de odio. Por eso la interpretación como mando disfrazado es una posibilidad por investigar y no una clasificación del enunciado, y por eso queda, en los recuadros del capítulo, en la columna de la hipótesis.

5.3 Control aversivo dirigido a un grupo

Skinner (1957) presenta el control aversivo de la conducta verbal por la reducción de la estimulación aversiva. Cuando se reduce el propio estímulo aversivo, la conducta es de escape; cuando se reduce una condición que suele precederlo, es de evitación. Sus ejemplos verbales son "¡Basta!", reforzado cuando hace cesar la lesión física, y "¡No me toques!", reforzado cuando hace cesar su amenaza. El punto que importa aquí es que la estimulación aversiva no es solo física. Skinner escribe que un hombre puede estimular a otro aversivamente de muchos modos, incluso por "muchas clases de 'daño verbal'" (p. 54), y que la amenaza que controla la conducta verbal no implica necesariamente violencia: hay formas leves, y una pregunta contiene una ligera amenaza generalizada, en el sentido de que, si no se responde, sigue la censura.

CONCEPTO EN USO

Control aversivo (Skinner, 1957; Fazzano y Gallo, 2015)

El control aversivo es el control de la conducta por estimulación aversiva: la conducta que la remueve o la evita es reforzada (escape y evitación), y la que la produce tiende a ser suprimida (castigo). Fazzano y Gallo (2015) definen los castigadores, con Sidman (1989/2003), como la producción de algo que normalmente sería un reforzador negativo o la retirada de algo que normalmente sería un reforzador positivo. En el caso que analizan, la aversividad no está en el estímulo: es aprendida. El recorte que hacen es el de una contingencia de tres términos, que reconocen como una simplificación de las contingencias implicadas en las agresiones (nota 3). La estimulación aversiva puede ser verbal (Skinner, 1957) y puede ser leve.

El artículo de Fazzano y Gallo (2015) es el análisis conductual brasileño de la homofobia que traen las fuentes verificadas, y es, por declaración propia, una hipótesis funcional y no un análisis funcional: se formula la hipótesis, y no se experimenta, por razones éticas. En sus términos, la homofobia es un conjunto de conductas complejas, operantes y respuestas emocionales, de agresión física, psicológica o sexual contra homosexuales o contra quien se identifique con la cultura homosexual, y se relaciona con el control aversivo. La parte que interesa a este capítulo es la que trata del mantenimiento. Para el agresor, las consecuencias inmediatas son la remoción de la estimulación aversiva, que las autoras suponen aprendida, y, tal vez, la aprobación del grupo. Las conse-

cuencias punitivas, como procesos penales, multas o prisión, son distantes en el tiempo, y el agresor puede no entrar nunca en contacto con ellas, porque, según las autoras, en 2015 no había agravantes para la homofobia en el código penal y el número de denuncias registradas era reducido. El texto es de 2015, y el marco jurídico puede haber cambiado; no se afirma aquí el estado actual de la ley. Al mismo tiempo, para que las consideremos consecuencias punitivas, es necesario observar la recurrencia de la relación conducta–consecuencia. Lo cual no ocurre.

El mismo análisis tiene un lado de la víctima, y es el lado menos trabajado por los estudios del discurso. Fazzano y Gallo (2015) tratan la agresión, desde el punto de vista de quien la recibe, como castigo de la conducta que se emitía: dos hombres agredidos por caminar abrazados probablemente no volverán a abrazarse en el mismo lugar o en contextos parecidos, y el castigo se generaliza, de modo que la persona puede dejar de ir al lugar de la agresión y, en algunos casos, de salir de casa, aislándose. Quedarse en casa, evitar relaciones o permanecer "en el armario" serían respuestas incompatibles, que previenen nueva estimulación agresiva. Es un análisis de los efectos de agresiones, y no de enunciados, y el capítulo lo trae porque el enunciado de odio, si es aversivo para el grupo destinatario, tiene ese lado: la consecuencia que un enunciado así produce en el grupo es una reducción de lo que el grupo hace en público. Este análisis corresponde al segundo caso señalado por Skinner, el de la aversividad. Si un mando disfrazado resulta en agresiones y en la supresión de las conductas de los agredidos, la supresión refuerza negativamente la conducta de emitir el mando disfrazado. Ninguna fuente verificada mide esto para enunciados.

Hay una salvedad de método de las mismas autoras que vale para todo el libro. La conducta se define como violencia o agresión por su función en determinados contextos: dos amigos que se tratan con un término generalmente peyorativo pueden estar expresando solo afinidad e intimidad, y el mismo término caracterizaría una agresión verbal si generara en la víctima tristeza, vergüenza o humillación. Esto concuerda con la regla de rótulo de este libro. La palabra no decide: decide lo que hace, con quién y en qué ocasión.

Dos lagunas completan el cuadro. La primera es la que Mizael (2018) documenta en una revisión bibliográfica. La autora buscó los términos "homossexualidade", "orientação sexual", "heterossexualidade" y "bissexualidade" (y sus equivalentes en inglés y en español) en 13 revistas de Análisis de la Conducta, tres brasileñas y 10 internacionales, sin límite de período, e incluyó solo los artículos en que la homosexualidad era tema central. Encontró tres, y una búsqueda complementaria en SciELO, en portugués, que añadió el término "análise do comportamento", trajo el cuarto. Uno de ellos es el de Faz-

zano y Gallo (2015), el único publicado en una revista de psicología no específica del área. En general, los cuatro analizaban la homosexualidad o la homofobia por las contingencias de reforzamiento y castigo a las que las personas pueden haber sido sometidas o examinaban posibles causas biológicas de la homosexualidad. Faltaban, según la autora, análisis funcionales de conductas consideradas prejuiciosas y el estudio de chistes y otras conductas potencialmente discriminatorias como prácticas reforzadas por la comunidad verbal. La segunda laguna es la que Fazzano y Gallo (2015) reconocen. Los datos brasileños que citan vienen de Disque 100 (línea telefónica federal de denuncias de violaciones de derechos humanos), canal de denuncias en que se basó la Secretaría de Derechos Humanos, en 2011, para caracterizar la violencia homofóbica: 6.809 relatos de violaciones contra lesbianas, gais, bisexuales y transexuales, que involucraban a 1.713 víctimas y 2.275 sospechosos. En el 61,9% de los casos las víctimas conocían al agresor, y las formas de violencia más frecuentes fueron la psicológica (42,5%), la discriminación (22,3%) y la violencia física (15,9%). Como son solo denuncias, las propias autoras escriben que los números pueden no expresar significativamente la realidad brasileña.

5.4 El rótulo del grupo y la transferencia de función aversiva

El Capítulo 4 mostró cómo una palabra adquiere función evaluativa sin contacto directo con el referente: por pertenecer a una clase de equivalencia o por entrar en relaciones de oposición y de comparación con otras (secciones 4.2, 4.3 y 4.6). Para el discurso sobre grupos, eso quiere decir que el enunciado no necesita un contacto previo directo del oyente con el grupo para que una función aversiva se transfiera al grupo; basta que el rótulo del grupo esté en una clase o en una relación con estímulos de función aversiva. Fazzano y Gallo (2015) plantean el punto en otros términos: si la presencia de homosexuales no es aversiva por sí misma, la aversividad fue aprendida, y el aprendizaje puede ocurrir por moldeamiento, por seguimiento de instrucciones y por imitación, y el seguimiento de instrucciones es un proceso verbal. Si una frase con autoclítico asertivo califica de forma peyorativa a un grupo (“¡Las mujeres votan mal!”), aunque el oyente no haya entrado en contacto de forma directa con estímulos aversivos derivados del voto de las mujeres, la posibilidad de que las mujeres voten adquiere función aversiva por transferencia de función o transitividad.

Mizael (2018) da un ejemplo de lo que la transferencia quiere decir para el prejuicio sexual, y lo da como ilustración teórica y propuesta de investigación, y no como dato: un niño puede aprender que “viado” (insulto homofóbico del portugués de Brasil) es algo malo sin saber el significado de la palabra; cuando crece, aprende que la palabra es equivalente a “gay” u “homosexual”, y los

atributos negativos pasan a quien se define o es percibido como gay. El mecanismo que el ejemplo describe es el que el Capítulo 4 mostró en el laboratorio con otros estímulos. Dixon et al. (2003), en el mismo sentido, analizan el prejuicio y los actos de terror como producto de la transformación de funciones de individuos o grupos con base en un contacto frecuentemente indirecto con pocos de sus miembros, como argumento y sin datos (Capítulo 4, sección 4.6).

Lo que esto añade al capítulo es una tercera posición para el hablante, además de la de objeto y la de agente de aversión: la de quien establece las relaciones. Un enunciado que asocia el rótulo de un grupo a términos de función aversiva, o que lo opone a un "nosotros" de función positiva, estaría haciendo, para el oyente, el trabajo que el procedimiento de laboratorio hace para el participante. Como dato, aquí, solo está el procedimiento; la hipótesis de que el discurso político lo haga, y con qué efecto sobre lo que el oyente dice y hace después, se propone aquí, y para ponerla a prueba serían necesarias declaraciones fechadas y medidas del oyente antes y después, y no solo el análisis del texto.

5.5 Del discurso al acto: lo que los datos permiten afirmar

El paso del discurso al acto junta dos preguntas que conviene separar. La primera es si el enunciado cambia lo que la audiencia hace, por ejemplo, actuar contra el grupo o dejar de castigar a quien actúa. La segunda es si el enunciado cambia lo que el propio hablante u otros hablantes dicen a continuación. Los estudios verificados son, en su mayor parte, de la segunda; los de la primera no tratan del discurso de odio. La Tabla 5.1 los reúne.

TABLA 5.1

Lo que seis estudios verificados miden como "discurso", como "acto" y lo que permiten afirmar

Estudio	Medida de "discurso"	Medida de "acto" o de efecto	Lo que permite	Lo que no permite
Ajzenman et al. (2023, versión de trabajo)	Participación del presidente en una manifestación (15/3/2020) y pronunciamiento (24/3/2020) contra el distanciamiento social; no es discurso sobre un grupo	Índice de distanciamiento por teléfonos móviles, por municipio	Caída del distanciamiento mayor en los municipios donde Bolsonaro obtuvo más del 50% de los votos en la primera vuelta de 2018, cerca de 1 punto porcentual en promedio, sin efectos antes de los eventos	Separar conducta de declaración; hablar de discurso de odio; el mecanismo (medios, evangélicos) es solo sugestivo

Estudio	Medida de "discurso"	Medida de "acto" o de efecto	Lo que permite	Lo que no permite
Bursztyn et al. (2020)	Información aleatoria sobre quién ganó en el área (popularidad percibida de Trump)	Donación a una organización antiinmigración, en público o en privado	La donación pública, y no la privada, varió con la popularidad percibida	Efecto de declaraciones; solo manipula la percepción de popularidad; EE. UU.
Álvarez-Benjumea y Winter (2018)	Comentarios en un foro experimental, con puntuación de hostilidad	El propio comentario	Remover el contenido de odio anterior redujo los comentarios hostiles (del 5,88% al 0,99%); el contradiscurso no tuvo efecto significativo	Generalizar; participantes pagados y anónimos
Siegel et al. (2021)	Tweets clasificados por diccionario y por comunidades de Reddit	Nivel agregado del discurso, sin medida de acto	Sin aumento persistente del discurso de odio en la campaña ni después de la elección de 2016	Excluir otras plataformas; diccionario conservador
Marino-Fages y Martínez-Pozo (2026, documento de trabajo)	Tweets geolocalizados clasificados por modelo y por diccionario	Nivel del discurso, por municipio	Aumento después de la elección de 2018, mayor donde Bolsonaro obtuvo menos votos, en homofobia y sexismo	Medir el costo social percibido; el mecanismo es lectura de los autores
Hangartner et al. (2021)	Tuits xenófobos de 1.350 usuarios	Eliminación retrospectiva y creación prospectiva de discurso de odio	El contradiscurso basado en la empatía aumentó la eliminación y redujo la creación; humor y aviso de consecuencias, sin efecto consistente	Efecto sobre quien solo lee

Nota. Números de los artículos. Ajzenman et al. se leyeron en la versión de trabajo de 2020, de la Universidad Torcuato Di Tella, sin la paginación de la revista. Marino-Fages y Martínez-Pozo es un documento de trabajo que, según la portada, no pasó por revisión de pares. Los estudios de Bursztyn et al., de Álvarez-Benjumea y Winter, de Siegel et al. y de Marino-Fages y Martínez-Pozo se describen en el Capítulo 10 (secciones 10.3, 10.4 y 10.6); el de Hangartner et al. se trata en el Capítulo 12.

El estudio de Ajzenman et al. (2023) es el único, entre los verificados, que vincula la declaración de un jefe de Estado brasileño con lo que hace su audiencia, y lo que mide es el distanciamiento durante la pandemia y no un acto contra un grupo. La hipótesis de los autores era que los partidarios del presidente serían mucho más propensos a dejarse persuadir por sus discursos que sus críticos, incluso en decisiones de alto costo, como seguir o no las medidas preventivas recomendadas por la Organización Mundial de la Salud y por los gobiernos subnacionales. El desenlace fue un índice de distanciamiento cal-

culado cada día, para cada municipio, a partir de la localización de cerca de 60 millones de teléfonos móviles anónimos: la proporción de los aparatos que permaneció, ese día, a menos de 450 metros del lugar identificado como la casa. El índice va de 0 a 1, tuvo una media de 0,37 en el período (de febrero a las dos primeras semanas de abril de 2020) y existe para 3.975 de los 5.571 municipios, porque los pequeños no tienen aparatos suficientes (sección II). Los dos eventos estudiados fueron la participación del presidente en las manifestaciones del 15 de marzo de 2020 y el pronunciamiento oficial del 24 de marzo, en el que presionó por el fin de las medidas de distanciamiento de los gobiernos subnacionales.

La comparación es entre los municipios progubernamentales, en los que Bolsonaro obtuvo más del 50% de los votos en la primera vuelta de 2018, y los demás, día a día, antes y después de los dos eventos. Los autores llaman al diseño estudio de eventos, con efectos fijos de día y de municipio (que descuentan lo que es propio de cada municipio y lo que cambia en todos el mismo día) y estimaciones separadas para cada día anterior (adelantos) y posterior (rezagos) al evento (secciones I y V.1). En los días anteriores, con una excepción, la diferencia entre los dos grupos no difiere de cero, es decir, no había una diferencia previa entre ellos. A partir del segundo día después de los eventos, el distanciamiento fue significativamente menor en los municipios progubernamentales, cerca de 1 punto porcentual en promedio. En una especificación que se restringe a los estados en los que obtuvo más del 50% de los votos en al menos un tercio de los municipios y compara, en cada estado, los municipios por encima y por debajo de la mediana de votos, el efecto queda cerca de 2 puntos porcentuales, frente a una media de 0,37 en el período (V.2). Con los coeficientes y la población de los municipios progubernamentales (cerca de 100 millones), los autores calculan, en una cuenta aproximada, cerca de 1 millón de personas adicionales alejadas más de 450 metros de su casa en cada uno de los diez días siguientes. Llaman al resultado un "fuerte efecto de persuasión", y dicen no poder presentar resultados concluyentes sobre los mecanismos (V.3 y VI).

En términos operantes, el estudio es compatible con una lectura del pronunciamiento como regla, o como mando, para quien ya tenía una historia de reforzamiento con el presidente, y es compatible también con la imitación de su conducta en la manifestación. Los autores estiman el efecto conjunto de la participación en las manifestaciones y de los mensajes contra el distanciamiento (sección V.1) y, por lo tanto, no distinguen entre los dos. El provecho para este capítulo es de método: medir lo que la audiencia hace después de una declaración es viable a gran escala para conductas que registran los teléfonos móviles. En cambio, la pregunta por el acto contra un grupo exigiría

otra medida. La condición de lectura es que el efecto aparece donde la audiencia ya apoyaba al líder. Baldwin y Baldwin (1986) sostienen que aspectos de la relación del observador con el modelo (admiración, respeto y afecto) son elementos fundamentales en la probabilidad de que el observador se comporte como el modelo. No es un dato sobre discurso de odio, y no debe leerse así.

Los cuatro estudios siguientes, descritos con más detenimiento en el Capítulo 10, tratan de lo que las personas dicen o aceptan hacer en público después de una información o de un evento, y dos de ellos son experimentos. Bursztyn et al. (2020) reclutaron a 1.600 residentes de la región de Pittsburgh en un panel en línea e informaron, al azar, a la mitad de ellos de que Trump había ganado en la región metropolitana y a la otra mitad de que Clinton había ganado en el condado donde está la ciudad. Ofrecieron entonces una bonificación a quien autorizara una donación a una organización fuertemente antiinmigración, y la decisión era anónima (condición privada) o visible para pares de la región (condición pública). En público, aceptaron donar el 33% de quienes supieron de la victoria de Trump y el 20% de quienes supieron de la de Clinton; en privado, el 31% y el 30%. La diferencia entre esas diferencias fue significativa: el efecto de la información apareció en público y no en privado.

Álvarez-Benjumea y Winter (2018) reclutaron a 180 participantes en una plataforma de crowdsourcing y les pidieron que comentaran hechos en un foro experimental, en cuatro condiciones: la línea base, con comentarios amistosos, neutros y hostiles; la censurada, con el contenido de odio anterior borrado; la extremadamente censurada, solo con comentarios amistosos; y la de contradiscurso, con los comentarios hostiles acompañados de réplicas que destacaban su inaceptabilidad. Los comentarios hostiles fueron el 5,88% en la línea base y el 0,99% en la condición censurada, y solo esa reducción fue significativa; el contradiscurso (4,44%) no tuvo efecto significativo.

Los otros dos son estudios observacionales de textos en redes sociales. Siegel et al. (2021) analizaron más de 750 millones de tweets sobre la elección estadounidense de 2016 y casi 400 millones de una muestra aleatoria de usuarios estadounidenses, y con las dos medidas que adoptaron (un diccionario ampliado por aprendizaje automático y una clasificación basada en comunidades de Reddit vinculadas a la derecha alternativa) no encontraron un aumento persistente del discurso de odio en la campaña ni después de la elección; hubo solo picos breves.

Marino-Fages y Martínez-Pozo (2026) reunieron más de 37,6 millones de tweets en portugués, geolocalizados en Brasil, de julio de 2017 a diciembre de 2019, y compararon el antes y el después de la elección de 2018 según la proporción de votos de Bolsonaro en la primera vuelta en cada municipio: el dis-

curso de odio aumentó después de la elección, más donde él obtuvo menos votos, y el aumento se concentró en homofobia y sexismo. Las cuatro medidas y los cuatro diseños difieren, y los estudios no se confirman ni se refutan entre sí. Hangartner et al. (2021), en el Capítulo 12, completan la lista del lado de la intervención: sortearon a usuarios de Twitter autores de tuits xenófobos para recibir o no un contradiscurso, y el de empatía aumentó la eliminación del tuit original y redujo la creación de nuevos tuits xenófobos, con salvedades que el Capítulo 12 detalla.

Álvarez-Benjumea y Winter (2018) sacan del experimento una conclusión que interesa a la lectura operante: la conducta normativa en línea estaría motivada más por normas descriptivas - lo que hacen los demás - que por normas prescriptivas - lo que se debe hacer. En el vocabulario usado aquí, la norma descriptiva es lo que la audiencia hace ante el enunciado, estímulo discriminativo para quien va a hablar a continuación; la prescriptiva es lo que la comunidad dice que se debe hacer, una regla (sección 3.7). La condición de contradiscurso trae la prescriptiva, con réplicas que dicen que el odio es inaceptable, y no tuvo efecto significativo; la condición censurada cambia lo que se ve decir a los demás y tuvo efecto. Los autores sugieren, para el fracaso del contradiscurso, que las réplicas, ellas mismas hostiles, añaden ambigüedad sobre la norma. El experimento no separa las dos lecturas y es un solo experimento, con participantes pagados y anónimos; la lectura es compatible con la de que, para quien habla, lo que la audiencia hace pesa más que lo que ella dice que se debe hacer.

Falta, en la Tabla 5.1, la columna que el título de la sección pedía: la del acto contra el grupo. Entre las fuentes verificadas, ninguna vincula enunciados de odio con actos de violencia contra el grupo destinatario, en Brasil o fuera de él. Hay estudios que lo hacen en la literatura, y no se citan aquí porque su texto completo no pudo leerse por vía legal. La laguna es, por lo tanto, del acervo de fuentes leídas y debe leerse como tal: no es ausencia de relación, ni su demostración.

DATO

En los municipios donde el presidente obtuvo más votos, el distanciamiento social cayó después de dos eventos en los que él se opuso públicamente al distanciamiento, cerca de 1 punto porcentual en promedio, sin efecto antes de los eventos (Ajzenman et al., 2023). La expresión pública de una posición antiinmigración, y no la privada, acompañó a la popularidad percibida del candidato (Bursztyn et al., 2020). En Brasil, el discurso de odio en tweets geolocalizados aumentó después de la elección de 2018, más donde Bolsonaro obtuvo menos votos (Marino-Fages y Martínez-Pozo, 2026).

INTERPRETACIÓN	Lo que la audiencia hace y lo que los demás dicen parecen depender de lo que la audiencia cree que los demás aceptan y de lo que el líder hace y dice. Las evidencias son indirectas, de diseños y medidas diferentes, y ninguna mide la declaración de un político de modo que pueda atribuirse el efecto al contenido y no a la conducta, al cargo o al momento.
HIPÓTESIS	Que el enunciado de odio de un político funcione como mando a la audiencia, con un efecto medible sobre lo que ella hace con el grupo destinatario, es una hipótesis, y ninguna fuente la pone a prueba. Ponerla a prueba exigiría declaraciones fechadas, una medida de lo que la audiencia hace después y la comparación con quien no estuvo expuesto y, para el grupo destinatario, medidas de lo que deja de hacer en público.

5.6 Un caso con clasificación externa: la declaración de 2011 y lo que vino después

El caso de la apertura es el único entre las fuentes verificadas con los tres elementos que la regla del capítulo pide para hablar de discurso de odio sin rotular por cuenta propia: el enunciado transcrito, la fecha y el medio, y la clasificación por una instancia externa, nombrada y fechada. Lo que no tiene es lo que la ficha del Capítulo 3 pide en audiencia presente y en variables de control candidatas, la composición de la audiencia y su reacción, que quedan en blanco; la sección dice cuáles.

Los campos documentados son estos. La declaración: “Eso ni me pasa por la cabeza, porque ellos tuvieron una buena educación. Soy un padre presente, así que no corro ese riesgo”, en respuesta a una pregunta sobre qué haría si tuviera un hijo gay (Agência Lupa, 2026). La fecha y el medio: 2011, programa CQC, de TV Bandeirantes; el mes no consta en la fuente. La clasificación externa: condena por daño moral colectivo, confirmada por el TJRJ (Tribunal de Justicia de Río de Janeiro) en noviembre de 2017, tras declaraciones de contenido homofóbico al programa, con una pena fijada en 2015 de R$ 150 mil, todo según Lupa; el texto del tribunal no se leyó. También según Lupa, el proceso seguía en trámite en 2018 y hay otras declaraciones del diputado con un contenido que ella llama homofóbico, en un debate en TV Câmara (el canal de la Cámara de Diputados) y en una entrevista a *Playboy*, en 2011. El 3 de agosto de 2018, en la misma verificación, la frase “Yo nunca fui homofóbico”, dicha en el programa Central das Eleições, de GloboNews, fue clasificada por Lupa como “Falso” (Capítulo 3, sección 3.8, caso 4).

Los campos de la ficha que quedan en blanco son los más importantes para el análisis operante: la reacción de la audiencia del programa, su composición, lo que el hablante hizo después con audiencias diferentes y cualquier medida de lo que el grupo destinatario hizo o dejó de hacer. Nada de eso está en las fuentes, y el capítulo no los llena por inferencia. La negación de 2018 fue clasificada por Lupa y figura entre los casos del Capítulo 3, donde recibió el mismo tratamiento que las declaraciones de Lula; el análisis allí es de la forma de la frase y de la correspondencia con el dato, y el de aquí, de la relación entre la declaración de 2011, la sanción y la declaración de 2018.

Una salvedad de simetría, en el mismo espíritu de la que hizo el Capítulo 3. Las fuentes verificadas no traen, para otro campo político, un caso equivalente, es decir, con enunciado transcrito, fechado y clasificado por una instancia externa nombrada. Eso es un límite del acervo de fuentes leídas y no una propiedad de los campos políticos. La ficha y el criterio serían los mismos, y la regla de muestra de la sección 2.9 es lo que permitiría encontrar los casos.

DATO	Una respuesta, dada en 2011 en un programa de televisión a una pregunta sobre qué haría el diputado si tuviera un hijo gay, llevó a una condena por daño moral colectivo confirmada en 2017, según Lupa. En 2018, el mismo político, en una entrevista, dijo que nunca había sido homofóbico, y Lupa clasificó la frase como "Falso".
INTERPRETACIÓN	La respuesta se emitió ante una pregunta, y el enunciado se apoya, en la forma, en dos descripciones del propio hablante ("soy un padre presente", "ellos tuvieron una buena educación") y en una palabra de función evaluativa ("riesgo"), que hace de tener un hijo gay una ocasión que evitar. Solo por la forma, no se decide si el enunciado es un tacto sobre lo que el hablante haría, un intraverbal aprendido o un mando a la audiencia (secciones 3.3 y 5.2). La condena es una consecuencia aversiva, tardía en años, y viene de una instancia jurídica y no de la audiencia del programa. La condena por daño moral colectivo es una clasificación jurídica del efecto sobre el grupo; no es una medida de la conducta del grupo destinatario.
HIPÓTESIS	Que la declaración de 2011 haya sido reforzada por la audiencia del programa (atención, aprobación, adhesión) es una hipótesis, y la fuente no registra nada sobre su reacción. Que la condena haya alterado la emisión de enunciados de ese tipo y que la negación de 2018 ("Yo nunca fui homofóbico") sea función de esa historia, ya como evitación del castigo, ya como tacto distorsionado por la consecuencia del oyente (sección 3.4), es una hipótesis y exigiría una serie de declaraciones fechadas, antes y después de 2015 y de 2017, con las audiencias.

5.7 Interpretaciones rivales y límites

La lectura del capítulo, de que el enunciado de odio se mantiene por consecuencias sociales y aversivas y puede funcionar como mando disfrazado, tiene rivales, y la Tabla 5.2 las reúne con lo que las separaría de ella.

Los límites del capítulo son los de las fuentes. Sobre la agresión y el discurso de odio, el acervo de fuentes leídas trae un artículo de hipótesis funcional sobre la homofobia (Fazzano & Gallo, 2015) y una revisión que muestra la escasez del área (Mizael, 2018); los experimentos del Capítulo 4 tratan del prejuicio, no de enunciados de odio, no hay estudio que ponga a prueba la función de mando o el control aversivo para enunciados. Las evidencias sobre el discurso y el acto son de diseños y medidas diversos, de un país, de una plataforma o de un foro experimental, y ninguna es de declaración de un político con

audiencia identificada. El único caso con clasificación externa es de un solo campo político, y la clasificación es de una instancia y de los relatos de Lupa, registrados con la atribución. Las relaciones discontinuas de la Figura 5.1 no tienen medida entre las fuentes. La sección 5.5 dice qué lagunas son del acervo de fuentes leídas, y no de la literatura.

TABLA 5.2

Interpretaciones rivales para un enunciado clasificado como de odio y lo que las separaría de la lectura operante principal

Rival	Lo que dice	Lo que la separaría
Estado emocional	El enunciado es efecto de un estado de ira o de aversión, más que de una consecuencia social; Skinner observa que provocar ira aumenta la probabilidad de conducta abusiva o agresiva (1957)	Si la emisión varía con la presencia, la ausencia y la identidad de la audiencia; si cesa cuando no hay nadie
Identidad y ritual de grupo	El intercambio es el reforzamiento, y el contenido varía sin alterar el efecto (Guerin, 1994)	Si el contenido cambia sin cambiar la reacción; si la reacción es a la adhesión y no a lo dicho
Intraverbal	Repite lo que se oyó en la familia, en la religión o en el grupo, por modelado y regla (Fazzano y Gallo, 2015)	Historia de adquisición; si varía con la fuente verbal; si es sensible o insensible a las consecuencias
Broma	La función es la afinidad o la atención y no la agresión (Fazzano & Gallo, 2015; Mizael, 2018)	El efecto sobre el grupo destinatario y sobre la audiencia; si el término genera vergüenza o humillación en quien es descrito
Tacto bajo control de un evento	La frase es el relato de un evento, con correspondencia verificable	Verificación de la correspondencia por una agencia; si cambia con el evento
Artefacto de la clasificación	Otra instancia, con otro criterio, no lo clasificaría así (los clasificadores y diccionarios usados en los estudios se equivocan: exactitud del 77 al 79% en Marino-Fages y Martínez-Pozo, 2026)	Comparar criterios e instancias; reclasificar una muestra por más de un criterio

Síntesis

Un enunciado clasificado como de odio es un rótulo de una instancia externa y no una explicación. El análisis operante pregunta qué consecuencias lo mantienen, qué función tiene para la audiencia y qué efecto tiene para el grupo al que se refiere. Para el hablante, las candidatas son la aprobación de la audiencia, la remoción de estimulación aversiva aprendida y funciones que

varían con la ocasión. Para la audiencia, la hipótesis es que un enunciado con forma de tacto opere como mando disfrazado: condición de aparición, una audiencia que castiga el mando explícito; criterio, forma de una cosa, función de otra. Para el grupo destinatario, la estimulación es aversiva y, en el análisis de Fazzano y Gallo (2015) para la agresión, castiga lo que el grupo hacía en público y refuerza lo que lo oculta.

Los datos verificados no vinculan el enunciado de odio con un acto contra el grupo. Muestran, con medidas y diseños que no se suman, que la declaración de líderes precede a cambios en lo que hacen sus partidarios (en el distanciamiento social, no en el odio), que la expresión pública acompaña a la popularidad percibida, que la remoción moderada de contenido hostil reduce la hostilidad en un foro experimental y que, en Brasil, el discurso de odio en tweets aumentó después de la elección de 2018, más donde el vencedor obtuvo menos votos. El único caso con clasificación externa en las fuentes es la declaración de 2011 y la condena confirmada en 2017, según Lupa; los campos de la ficha que importan al análisis operante están en blanco, y así quedan.

El Capítulo 6 retoma el control aversivo desde el lado de quien lo recibe, como contracontrol (sección 6.4); el 10 define la ventana de una audiencia por lo que castiga y por lo que no castiga y describe con más detenimiento los estudios de la Tabla 5.1; el 12 trata del contradiscurso y de las demás intervenciones sobre el hablante y sobre la audiencia.

Referencias

Agência Lupa. (2026, 27 de julio). *Anatomia da desinformação eleitoral* [Informe]. 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., 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), Artículo 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. En Associação Brasileira de Psicologia e Medicina Comportamental, *Comportamento em foco* (Vol. 2, pp. 157–170). 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, Trads.; 10.ª ed.). Martins Fontes. (Obra original publicada en 1953)

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

instituto WALDEN4 | ACADEMIA WALDEN4



PRÁCTICA BASADA EN LA EVIDENCIA

Cómo una práctica se convierte en evidencia, y lo que la evidencia no dice

El módulo completo sobre práctica basada en la evidencia: diseños, criterios, niveles de evidencia y los límites de cada uno. La misma pregunta que este libro les hace a los datos, hecha a la clínica.

Acceda, gratis y sin registro

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

Academia Walden4 · abierta, gratuita y sin registro

CAPÍTULO

6

Burbujas digitales como comunidades verbales y la evitación del contracontrol

Nyhan et al. (2023) pusieron a prueba, en un experimento de campo en Facebook durante la elección presidencial estadounidense de 2020, lo que ocurre con las actitudes cuando se reduce la exposición a fuentes de la misma inclinación política. Participaron 23.377 usuarios en los Estados Unidos. Para los asignados al tratamiento, los autores relegaron en el feed, del 24 de septiembre al 23 de diciembre de 2020, el contenido de amigos, grupos y páginas que se preveía que compartían la inclinación política del participante; el feed del grupo de control no se alteró de modo sistemático. La proporción de lo que los usuarios veían procedente de esas fuentes cayó del 53,7% (control) al 36,2% (tratamiento), cerca de un tercio menos. Los autores midieron entonces ocho actitudes que habían registrado de antemano (prerregistro), como polarización afectiva, extremismo ideológico, evaluación de candidatos y creencia en afirmaciones falsas, y no encontraron efecto medible en ninguna: descartan efectos de 0,12 desviaciones estándar o más, es decir, de poco más de una décima parte de la variación típica de cada medida. Los datos de exposición del mismo artículo, de 2020, que abarcan a la población adulta activa de Facebook en los Estados Unidos, indican que el contenido de fuentes de la misma inclinación es la mayor parte de lo que los usuarios ven. Si la exposición homogénea fuera lo que sostiene las actitudes, recortar un tercio de ella debería mover algo, y no lo movió.

El resultado admite dos lecturas. En la primera, la burbuja no tiene el efecto que se le atribuye. En la segunda, lo que se redujo no era lo que importa: la literatura mide la burbuja por la exposición, es decir, por lo que llega al feed, y el análisis de la conducta verbal pregunta por otra cosa, por quién refuerza lo que dice el hablante y por cuánto de lo que viene de fuera de la comunidad

que lo refuerza llega hasta él. Este capítulo toma la segunda lectura como hipótesis de trabajo y dice, en cada paso, qué la sostiene. La primera se presentará, con sus argumentos, en la sección 6.8.

El recorrido tiene tres partes. Las secciones 6.1 a 6.3 describen la burbuja como comunidad verbal y la confrontan con los datos sobre cámaras de eco. La sección 6.4 trata de lo que protege al hablante de la corrección, que es la propuesta central del capítulo y la parte con menor respaldo empírico. Las secciones 6.5 a 6.7 lo llevan al material brasileño, incluido un análisis exploratorio de dos videos políticos en Instagram (sección 6.6), y a un método del autor. En todas, dato, interpretación e hipótesis se mantienen separados, y llamar evitación del contracontrol a un repertorio es siempre interpretación.

6.1 La burbuja como comunidad verbal: audiencia y fuente de reforzamiento

Guerin (1992) recuerda que la comunidad verbal de Skinner, a menor escala, equivale a los colectivos de Durkheim y a los otros generalizados de Mead, y que los cambios en el uso del lenguaje pueden estar bajo el control de subcomunidades verbales más pequeñas. En ese sentido habla este capítulo de burbuja: no un lugar ni una distribución de contenidos, sino un conjunto de oyentes con una historia de reforzamiento común, que refuerza ciertas cosas dichas y otras no.

CONCEPTO EN USO

Subcomunidad verbal (Guerin, 1992), retomada del Capítulo 3

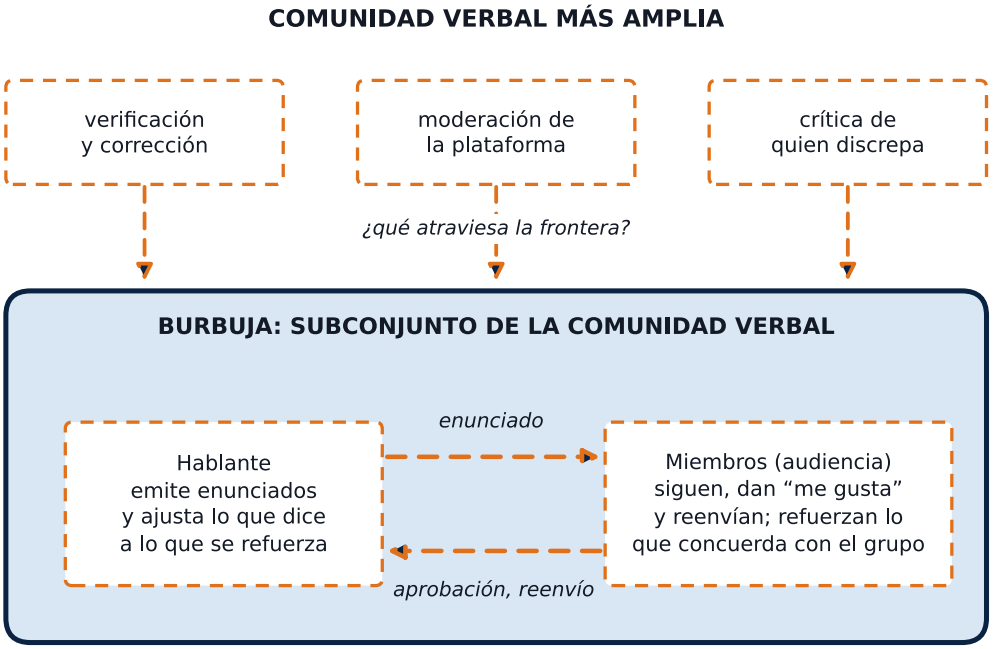
El Capítulo 3 (sección 3.4) presentó el tacto distorsionado: el tacto cuyo control por el evento ha sido desplazado por consecuencias especiales del oyente (Skinner, 1957). Lo que este capítulo añade es la escala. Para Guerin (1992), el control de lo que se dice puede estar en manos de subcomunidades verbales más pequeñas que la comunidad verbal amplia: si un grupo pequeño refuerza ciertos relatos y castiga otros, las características del grupo pasan a controlar lo que se relata, aunque las consecuencias sean generalizadas, y Skinner (1957, citado por Guerin, 1992) llamó distorsionados a esos tactos. La sección 6.3 presenta las tres vías por las cuales, según Guerin, lo que se dice se desliga del ambiente físico y pasa a ser controlado por el ambiente social.

El Capítulo 3 presentó los dos términos que la descripción necesita: la audiencia es ocasión para lo que se dice (sección 3.1), y su reacción es consecuencia (sección 3.3; Guerin, 1994). Falta decir qué cambia cuando la audiencia es una subcomunidad. Cambia la composición de las reacciones. Si los oyentes que reaccionan comparten una historia, las reacciones que recibe el hablante tien-

den a venir de la misma dirección, y lo que la comunidad más amplia diría del enunciado, corregir, criticar, no adherirse, tiene poco efecto o llega por otras vías. La hipótesis del capítulo es que eso, y no la mera exposición, es lo que hace de la burbuja un ambiente de control verbal distinto.

FIGURA 6.1

La burbuja como subcomunidad verbal: lo que miden los estudios y lo que supone el modelo



Línea continua: relación con datos (homofilia de las interacciones, Cinelli et al., 2021; fuentes de la misma inclinación en el feed, Nyhan et al., 2023). Discontinua: hipótesis sobre quién refuerza qué y cuánto de la corrección externa llega al hablante; ningún estudio del capítulo lo mide.

Nota. Esquema original del autor, sin datos. Línea continua: relación con datos (homofilia de las interacciones, Cinelli et al., 2021; predominio de fuentes de la misma inclinación en el feed, Nyhan et al., 2023). Discontinua: hipótesis. Las relaciones internas de la burbuja (quién refuerza qué) y lo que de la corrección externa llega al hablante no se midieron en ningún estudio del capítulo.

La literatura muestra homofilia en las interacciones y predominio de fuentes de la misma inclinación en el feed (sección 6.2). Las flechas internas, en las que el hablante ajusta lo que dice a lo que se refuerza y la audiencia refuerza lo que concuerda con el grupo, tienen apoyo en el análisis de Guerin y, para las opiniones, en los experimentos de 1956 a 1969 del Capítulo 9, pero no se midieron en burbujas digitales. Y la pregunta que la figura plantea en la frontera, cuánto de la corrección dirigida a lo que dijo el hablante llega hasta él, tampoco se midió: Nyhan et al. (2023) miden la proporción de contenido de fuentes transversales en el feed, que no es lo mismo.

6.2 Lo que dicen los datos sobre las cámaras de eco (y dónde divergen)

La literatura usa "cámara de eco" de modos diferentes, y la elección de la medida influye en la respuesta observada. Cinelli et al. (2021) definen la cámara de eco como un ambiente en el que la opinión, la inclinación política o la creencia del usuario sobre un tema se ve reforzada por interacciones repetidas con pares o fuentes de tendencias y actitudes semejantes, y la operacionalizan mediante dos datos observables, la homofilia en las interacciones sobre un tema (usuarios que interactúan con quienes piensan como ellos) y el sesgo de la difusión de la información a partir de pares de la misma inclinación, que deben cuantificarse de forma independiente. Analizaron más de 100 millones de contenidos sobre temas controvertidos (como el control de armas, la vacunación y el aborto) en cuatro plataformas, Gab, Facebook, Reddit y Twitter, con un conjunto de datos por plataforma: aborto (Twitter), vacunas (Facebook), política (Reddit) y todo Gab. La inclinación de los usuarios, un valor de −1 a +1 sobre el tema, se infirió de los "me gusta" en páginas rotuladas (en Facebook, por ejemplo, antivacuna vale +1 y provacuna, −1), de los enlaces a medios clasificados por Media Bias/Fact Check (2.190 medios) y de los comentarios. La homofilia se midió por la correlación entre la inclinación de cada usuario y la media de la inclinación de sus vecinos en la red de interacciones, y el sesgo de la difusión, por la inclinación media de los usuarios alcanzados a partir de usuarios de cada inclinación, en una simulación con un modelo epidémico (SIR, retomado en el Capítulo 11). Contrastan plataformas en las que el usuario puede ajustar el feed (Reddit) con las que no ofrecen esa opción (Facebook y Twitter) y atribuyen a la escasa moderación de Gab lo que describen como discurso de odio generalizado. En Facebook (vacunas) y en Twitter (aborto), la inclinación del usuario se correlacionó fuertemente con la de sus vecinos y la información llegó sobre todo a quienes tenían una inclinación parecida; en Reddit y en Gab, los usuarios formaron una única comunidad, inclinada a la izquierda (Reddit) o a la derecha (Gab). Como las diferencias entre plataformas podían reflejar los temas, los autores compararon Facebook y Reddit en un tema común, el consumo de noticias, y la segregación fue mayor en Facebook. La Tabla 6.1 resume el material y los resultados de los tres estudios que siguen.

Nyhan et al. (2023) miden otra cosa: lo que llega al feed de toda la población adulta activa de Facebook en los Estados Unidos. Solo el 20,6% de los usuarios recibe más del 75% del contenido de fuentes de la misma inclinación, y solo el 32,2% tiene una cuarta parte o más procedente de fuentes transversales; según los autores, los resultados no son consistentes con los peores temores so-

bre las cámaras de eco. En el experimento, la exposición a fuentes transversales subió del 20,7% al 27,9%. Quienes tuvieron la exposición reducida interactuaron más, por exposición, con el contenido de la misma inclinación que aún veían, y los autores dicen que los cambios probados en el algoritmo no anulan la propensión de los usuarios a buscar información afín e interactuar con ella. Para el efecto nulo citan, entre otras explicaciones, que el contenido político afín es una fracción de lo que se ve, que Facebook es una fracción de la dieta informativa y que el diseño no capta efectos del uso previo ni efectos acumulados a lo largo de años. Observan también que reducir la exposición podría llevar a las personas a buscar esa información en otro lugar de internet.

Bail et al. (2018), por su parte, hicieron el experimento en sentido inverso: aumentaron la exposición a opiniones opuestas. Una empresa de encuestas reclutó a republicanos y demócratas autodeclarados que visitaban Twitter al menos tres veces por semana; 1.652 respondieron a un cuestionario a mediados de octubre de 2017 (901 demócratas y 751 republicanos), con 10 afirmaciones sobre políticas públicas con las que debían declararse de acuerdo o en desacuerdo en una escala de siete puntos, que mide la posición liberal–conservadora. Una semana después, los investigadores sortearon quién recibiría la oferta de 11 US$ para seguir durante un mes a un bot, una cuenta automática de Twitter, que retuitearía 24 mensajes por día, sin que el participante conociera el contenido. Los bots, uno liberal y uno conservador, retuiteaban mensajes sorteados de una lista de 4.176 cuentas políticas (autoridades electas, líderes de opinión, medios de comunicación y organizaciones sin fines de lucro). Aceptaron la oferta el 64,9% de los demócratas y el 57,2% de los republicanos, y el mismo cuestionario se aplicó de nuevo cerca de un mes y medio después del primero. Los republicanos que siguieron a un bot liberal expresaron opiniones más conservadoras: el aumento fue de 0,12 puntos en la escala según la estimación más prudente, que considera a todos los sorteados para el tratamiento, hayan seguido al bot o no, y de 0,60 entre los plenamente adherentes, los que respondieron correctamente a todas las preguntas semanales sobre el contenido de los tuits del bot. Entre los demócratas, los efectos fueron pequeños, en dirección liberal, y no significativos. Los autores advierten que no identificaron el mecanismo, que a los participantes se les pagó por leer mensajes de quienes piensan distinto y que los hallazgos no deben leerse como evidencia de que la exposición a visiones opuestas aumente la polarización en todos los contextos. Su segunda hipótesis prerregistrada es que las personas expuestas a mensajes que entran en conflicto con sus actitudes tienden a contraargumentarlos con razonamiento motivado, lo que acentúa las diferencias percibidas y el compromiso con las creencias previas.

TABLA 6.1

Tres estudios sobre cámaras de eco: qué midieron, qué observaron y qué no permiten decir

Estudio	Material	Qué midieron	Qué observaron	Salvedad
Cinelli et al. (2021)	Más de 100 millones de contenidos de Gab, Facebook, Reddit y Twitter; inclinación inferida de "me gusta", enlaces y comentarios	Homofilia en las interacciones sobre un tema y sesgo de la difusión a partir de pares de la misma inclinación	La agregación en grupos homofílicos domina la dinámica; más segregación en Facebook (vacunas) y en Twitter (aborto) que en Reddit y en Gab, donde los usuarios forman una única comunidad	Mide interacción y difusión, no el efecto sobre lo que el usuario dice o declara (añadido); la inclinación es inferida
Nyhan et al. (2023)	Datos de 2020 de toda la población adulta activa de Facebook en EE. UU.; experimento de tres meses con 23.377 usuarios	Proporción de contenido de fuentes de la misma inclinación y de fuentes transversales en el feed; ocho medidas de actitud prerregistradas	Mediana del 50,4% (misma inclinación) y del 14,7% (transversales); exposición a fuentes de la misma inclinación del 53,7% al 36,2%, sin efecto medible en las actitudes	El contenido político afín es una fracción de lo que se ve; Facebook es una fracción de la dieta; el diseño no capta efectos del uso previo ni acumulados a lo largo de años
Bail et al. (2018)	1.652 encuestados (901 demócratas, 751 republicanos) de Twitter; 11 US$ por seguir durante un mes a un bot que retuiteaba 24 mensajes por día	Opiniones políticas en una escala de siete puntos después del mes	Republicanos tratados: aumento de 0,12 puntos (estimación más prudente) y de 0,60 (plenamente adherentes), en dirección conservadora; demócratas: efectos pequeños, en dirección liberal, no significativos	Mecanismo no identificado; participantes pagados por leer mensajes de quienes piensan distinto; no vale para todos los contextos

Nota. Los números proceden de los artículos. Cinelli et al. (2021) y Bail et al. (2018) se leyeron en XML de PMC, sin paginación; las páginas de Nyhan et al. (2023) son las de Nature. "(añadido)" marca una salvedad que no es de los autores.

Los tres trabajos parecen divergir, y en parte divergen, pero miden cosas diferentes: con quién se interactúa sobre un tema (Cinelli et al., 2021), lo que llega al feed entero, en todos los asuntos (Nyhan et al., 2023) y lo que declara quien es obligado a ver el otro lado (Bail et al., 2018). Es posible, sin contradicción, que un usuario interactúe de modo homofílico sobre un tema, reciba un feed poco homogéneo en conjunto y se desplace, al ver el otro lado, hacia donde ya estaba. Esto es interpretación, y no hallazgo de ninguno de los tres.

Lo que tienen en común importa más: ninguno mide las consecuencias que la audiencia dispone para lo que dice el hablante. Exposición, interacción y participación dicen qué llega y con quién se habla, y no quién refuerza qué. El dato más cercano es el de Nyhan et al. sobre la interacción: quienes vieron menos contenido afín interactuaron más, por exposición, con lo que aún veían. Cabe la lectura de que la propensión a buscar lo afín es función de una historia de consecuencias sociales y no de lo que ofrece el feed, y cabe también la hipótesis rival, de que al usuario simplemente le gusta lo que le gusta. El estudio no distingue entre las dos.

DATO	Hay una homofilia fuerte en las interacciones sobre temas específicos en Facebook y en Twitter (Cinelli et al., 2021); el feed del usuario mediano de Facebook tiene la mitad de fuentes de la misma inclinación, solo una quinta parte de los usuarios supera el 75% y, reducida la exposición afín en un tercio, no hubo efecto medible en las actitudes (Nyhan et al., 2023); la exposición forzada a mensajes del campo opuesto desplazó a los republicanos en dirección conservadora y no desplazó a los demócratas de modo significativo (Bail et al., 2018).
INTERPRETACIÓN	Los tres estudios miden lo que llega, con quién se interactúa y, en dos de ellos, lo que se declara después de cambiar la exposición. No miden las consecuencias que la audiencia dispone para lo que dice el hablante ni lo que de la corrección llega hasta él, de modo que no deciden si la burbuja opera como comunidad verbal.
HIPÓTESIS	Si la burbuja opera como subcomunidad verbal, el efecto sobre lo que se dice depende de quién refuerza y de cuánto de lo que viene de fuera llega al hablante, y no solo de lo que este ve. Ningún estudio del capítulo lo pone a prueba.

6.3 “Realidades virtuales”: cuando el control verbal sustituye al contacto con el evento

Esta sección no tiene datos: es el análisis teórico de Guerin (1992) y la extrapolación que se hace aquí. Guerin describe tres vías por las cuales el conocimiento socialmente construido se desliga del ambiente. En la primera, el control social del tacto pasa a un grupo que no refuerza el relato correcto del ambiente. En la segunda, intraverbales de alcance específico son reforzados por el grupo como si fueran tactos recordados. En la tercera, las consecuencias funcionales generalizadas se vuelven demasiado generalizadas y mantienen el tacto de modo indiscriminado. Guerin ilustra la segunda vía con enunciados como “los reactores nucleares son inseguros”: la mayoría de las personas no tiene experiencia directa con esos eventos, y los enunciados se mantienen por consecuencias sociales. Así, en condiciones socioeconómicas que favorezcan el uso de combustibles fósiles (contexto político favorable, precios internacionales bajos, conflictos geopolíticos, etc.), habrá una mayor propensión al reforzamiento de esas afirmaciones, a pesar de cualquier información fáctica adicional sobre la seguridad de los reactores nucleares. En 1994 retoma el ejemplo al tratar de actitudes sobre eventos que la persona nunca ha experimentado, y que por eso son intraverbales.

Un punto de Guerin merece decirse con todas las letras, porque descarta una lectura errónea: aun cuando lo que se dice es contrafáctico, las contingencias involucradas son reales, aunque sociales (1992). Quien afirma que hay un elefante rosa en la sala está bajo el control de contingencias sociales, y no del elefante. La expresión “realidades virtuales”, del índice de este libro, es un atajo para esas vías, y ninguno de los pasajes verificados de Guerin la usa; “virtual” no quiere decir aquí digital, y el artículo es anterior a las redes sociales.

Dos pasajes importan para el ambiente digital. En el primero, Guerin escribe que los medios de comunicación de masas pueden sostener representaciones sociales contrafácticas aun sin un grupo social, y que existe la posibilidad de que haya comunidades verbales reforzadoras solo en los medios, y no presencialmente. En el segundo, define el primer límite del conocimiento socialmente construido: la medida en que la conducta se mantiene por control de estímulos no social. La consecuencia que se extrae del primer límite es una extrapolación: donde el hablante tiene contacto directo con el evento, el relato queda bajo el control de este, y la distancia entre lo que el grupo refuerza y lo que el evento muestra tiende a hacerse visible. Guerin discute los límites tam-

bién mediante la distinción entre conducta gobernada por reglas y por contingencias (1992); el Capítulo 3 trata de ella. El segundo límite es el asunto de la sección siguiente.

6.4 Contracontrol y evitación del contracontrol

Guerin (1992) señala un límite a la medida en que la conducta puede ser mantenida por una comunidad verbal: una fuerte presión de conformidad produce, en general, contracontrol cuando especifica contingencias contrafácticas, a menos que se restrinja a un grupo pequeño y cohesionado, lo que solo funciona cuando los beneficios de pertenecer al grupo superan los de relatar con exactitud. El pasaje pide un término y una pregunta. El término es contracontrol. La pregunta es qué puede hacer un grupo para que no se formen las condiciones del contracontrol.

CONCEPTO EN USO

Contracontrol (Delprato, 2002)

Delprato (2002) define el contracontrol como episodios conductuales compuestos por condiciones controladoras aversivas socialmente mediadas y por respuestas de escape o evitación que no refuerzan la conducta controladora. No es un principio básico de la conducta: es siempre escape o evitación, y solo es singular porque el sujeto, ante una estimulación social o interpersonal aversiva, responde para oponerse al control en lugar de reforzarlo cediendo. Entre sus formas están salir del alcance del controlador, atacar y resistir pasivamente y, entre muchas otras, el absentismo, la crítica, el sabotaje, el incumplimiento de pedidos, la desatención y el abandono. Ocurre en dos niveles: escape o evitación de consecuencias específicas a corto plazo, o de contingencias coercitivas generalizadas y duraderas. Delprato asocia el contracontrol eficaz a condiciones del control: este debe ser conspicuo y estar equilibrado con el contracontrol, las fuentes deben estar identificadas y su conducta, accesible a los controlados; el control disfrazado y la delegación del control deben evitarse, porque vuelven inaccesible la conducta del controlador.

Dos puntos del texto importan para lo que sigue. El primero es que el contracontrol es respuesta al control aversivo. Delprato (2002) escribe que las personas valoran la libertad porque se refiere a métodos de control no aversivos, que por eso no ocasionan contracontrol, y que quienes tienen la conducta bajo reforzamiento positivo dicen sentirse libres porque las variables controladoras pueden ser inconspicuas. Para la burbuja, si la hipótesis de la sección 6.1 es correcta, la consecuencia es que el control ejercido sobre el miembro, hecho en buena parte de aprobación, tiende a no ocasionar contracontrol dentro del grupo, como en la pequeña comunidad cohesionada de Guerin. El con-

tracontrol que interesa viene de fuera o de los disidentes: la corrección de una agencia, la crítica de quien discrepa, la moderación de una plataforma, la salida de un miembro. El segundo punto es la lista de condiciones citadas por el autor y examinadas a continuación, que el capítulo usa como vara de medir.

Queda además una cuestión de sentido. En Delprato, el contracontrol es la respuesta de quien es controlado. La evitación del contracontrol, tal como la usa este capítulo, es el otro lado de la relación: lo que el hablante, o la comunidad que lo refuerza, hace que reduzca la probabilidad de que esa respuesta ocurra o su eficacia. Es una extensión propuesta aquí, y el recuadro de abajo lo dice sin rodeos.

CONCEPTO EN USO

Evitación del contracontrol (propuesta aquí, a partir de Delprato, 2002)

Repertorio del hablante, o de la comunidad que lo refuerza, mantenido por reducir la probabilidad o la eficacia del contracontrol: la corrección, la crítica, la denuncia, la moderación, la salida. El término no es de Delprato ni de otra fuente leída. Llamar evitación a un repertorio exige mostrar la contingencia: que la corrección es aversiva para el hablante o para la comunidad, que el repertorio la precede o la evita y que se mantiene por eso, y no solo por la aprobación de la audiencia (sección 6.8). Se tratan cuatro repertorios: forma, fuente, lugar y personas.

Forma. Guerin (1994) escribe que, si hay consecuencias negativas para ciertas declaraciones de actitud, estas pueden evitarse de varios modos por medio de autoclíticos. Relatar actitudes se refuerza cuando se presentan creencias de apoyo, pero solo si la comunidad verbal está de acuerdo con ellas; ante una audiencia que no cree en ellas, la creencia se omitiría. El Capítulo 2 (sección 2.8) argumentó que un enunciado que pasa de la forma fáctica a la opinativa deja de ser verificable, y Guerin muestra que los enunciados de actitud pueden funcionar como mandos. La lectura como evitación es que la forma le quita conspicuidad al control: el enunciado que actúa sobre el oyente sale como opinión. El dato de Guerin, sin embargo, tiene a la audiencia como variable, lo que sugiere que la protección de la forma es más necesaria en los márgenes que dentro de la burbuja. A un adepto incondicional de su grupo, el hablante le dice “las vacunas no funcionan”, mientras que a las personas que están en la frontera de su grupo les dice “no creo que las vacunas funcionen”.

Fuente. Evangelista y Bruno (2019) hicieron observación participante en al menos tres grupos políticos de WhatsApp y dos de Telegram, hasta el día de la elección de 2018 y en las semanas anteriores. En los grupos claramente identificados con Bolsonaro, observaron que se repetía que no había que creer en

los medios tradicionales, controlados por comunistas e izquierdistas, y que solo algunos youtubers, periodistas, blogueros y políticos merecían confianza. Después de la elección, el contenido de los medios tradicionales crítico con el nuevo gobierno se volvió menos común en los grupos, incluso cuando procedía de comentaristas considerados de derecha, mientras que antes el material crítico con el PT (Partido dos Trabalhadores, el partido de Lula) se compartía. Para los autores, la estrategia parece ser legitimar a los grupos como fuentes más fiables que los medios tradicionales. En la pandemia, Bonow Soares et al. (2021), que analizaron los mensajes más compartidos de más de 500 grupos políticos públicos de WhatsApp en marzo y abril de 2020 (802 mensajes con desinformación sobre la covid-19), dicen que el argumento político se usó para encuadrar como creados por la oposición todos los discursos que desacreditaban a Bolsonaro. Ricard y Medeiros (2020) registran, según la comisión parlamentaria mixta de investigación (CPMI) del Congreso Nacional que desde 2019 investiga la desinformación, que parte de la desinformación atacaba a gobernadores, medios y especialistas en salud para deslegitimar a quienes apoyaban el aislamiento. El cambio dentro del mismo tipo de fuente, compartida o no según a quién critica, es el dato más cercano al control por la audiencia: lo que se comparte varía con el blanco de la crítica, y no con la historia de la fuente.

Lugar. En el informe de 2026 de la Agência Lupa (agencia brasileña de verificación de datos), las verificaciones electorales son las de contenidos virales o potencialmente virales, sin autoría identificada, en los ciclos electorales de 2018 a 2024 (286 en total). Facebook aparecía en el 76,5% de las verificaciones de 2018, WhatsApp en el 92,4% de las de 2022 y en el 100% de las verificaciones electorales analizadas en 2024 (Agência Lupa, 2026). La agencia lee el dato como una migración de los contenidos engañosos, a medida que las redes abiertas ampliaron la moderación y pasaron a estar sometidas a mayor escrutinio público, hacia aplicaciones de mensajería y ambientes semiprivados, donde la autoría es más difícil de identificar y el monitoreo enfrenta mayores limitaciones. Recuero et al. (2025), que analizaron 811 verificaciones de tres agencias brasileñas sobre las elecciones de 2018 y 2022, leen el paso de Facebook a WhatsApp y Twitter, entre 2018 y 2022, como una adaptación a espacios más cerrados, en los que es más difícil que circulen contradiscursos, y Bonow Soares et al. (2021) escriben que WhatsApp es un canal clave de difusión de información falsa porque no puede ser rastreado ni desmentido públicamente por los verificadores. Los denominadores son pequeños (51 verificaciones en 2018, 131 en 2022 y 54 en 2024, leídos en el gráfico), y la propia Lupa advierte que mantiene una tipline en WhatsApp e integra programas de verificación de Meta, de modo que los resultados sobre plataformas son un retrato del universo que ella monitorea.

Personas. En los grupos que Evangelista y Bruno (2019) observaron, los comentarios críticos llevaban a la expulsión del autor y a la acusación de ser un petista (simpatizante del PT) infiltrado; los teléfonos de los acusados circulaban y se avisaba a los moderadores para que los excluyeran. Había poca discusión, y pocos miembros eran responsables de casi todo el contenido enviado o reenviado. Los autores señalan la invisibilidad de los actores que producen, monitorean, distribuyen o dirigen los contenidos: la mayoría de los usuarios desconoce la autoría de lo que recibe y no puede verificarla. Lupa escribe que la ausencia de autoría identificable de los ataques a candidatos dificulta la atribución de responsabilidades y favorece la rápida diseminación de esos contenidos (Agência Lupa, 2026). Los números sobre la concentración del origen de las imágenes son de Resende et al. (2019, citados por Evangelista y Bruno, 2019). Esos investigadores monitorearon grupos públicos de WhatsApp elegidos por la correspondencia entre la dirección de invitación y un diccionario de nombres de políticos, de partidos y de palabras asociadas al extremismo político: 141 grupos durante la huelga de los camioneros y 364 durante la elección. Un método automático identificó las imágenes compartidas que ya habían sido refutadas por servicios de verificación, y a ellas se sumaron 15 que Lupa había identificado como desinformación. Esas 85 imágenes con desinformación se compartieron ocho veces más que otras 69.590, verdaderas o no denunciadas a ninguna agencia independiente; eran el 1% de las imágenes y aparecieron en el 44% de los grupos monitoreados durante la campaña, y un pequeño número de grupos parecía originar gran parte de ellas.

TABLA 6.2

Cuatro repertorios de evitación del contracontrol: la condición de Delprato que cada uno debilita y las fuentes que lo describen

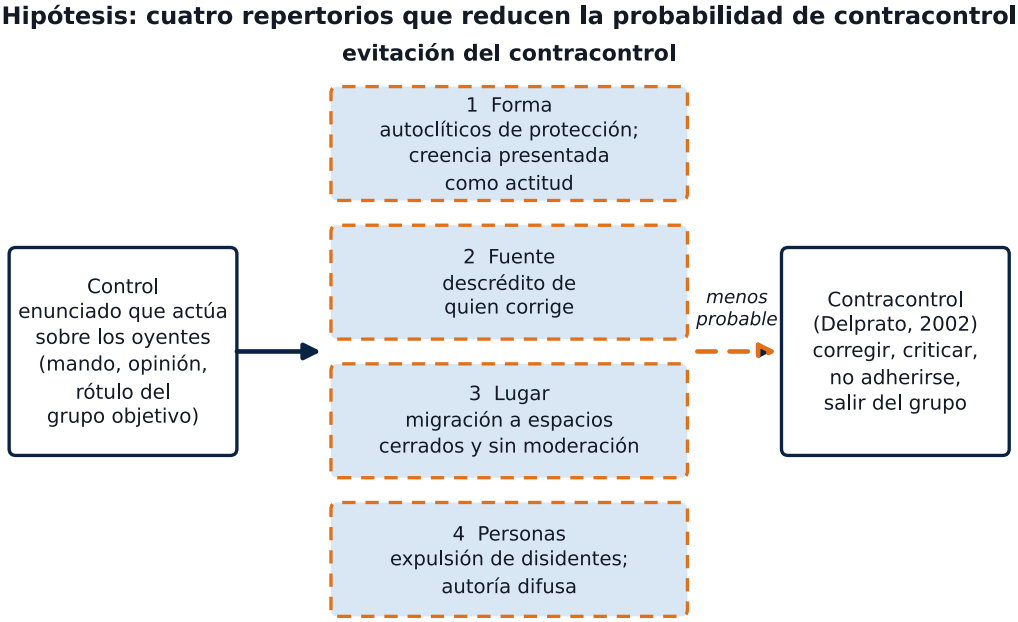
Repertorio	Condición de contracontrol eficaz que debilita (Delprato, 2002)	Fuentes que lo describen
1 Forma	Control conspicuo: el enunciado que actúa sobre el oyente sale como opinión o actitud	Guerin (1994)
2 Fuente	Ninguna: no figura en la lista de Delprato; añadido propuesto aquí	Evangelista y Bruno (2019); Bonow Soares et al. (2021); Ricard y Medeiros (2020)
3 Lugar	Conducta del controlador accesible a los controlados	Agência Lupa (2026); Recuero et al. (2025); Bonow Soares et al. (2021)
4 Personas	Fuentes identificadas; sin delegación del control	Evangelista y Bruno (2019); Agência Lupa (2026)

Nota. Páginas del PDF de las revistas y del informe de Lupa. La clasificación como repertorio de evitación es una interpretación propuesta aquí; las salvedades de cada fuente están en el texto de la sección.

La figura resume la hipótesis y no la prueba. Los repertorios 1, 3 y 4 debilitan condiciones de la lista de Delprato: la forma vuelve el control menos conspicuo, el lugar vuelve la conducta del controlador menos accesible y la autoría difusa vuelve la fuente menos identificable, y la expulsión de disidentes es la eliminación de quien ejercería el contracontrol. El repertorio 2 no está en la lista; actúa sobre la eficacia del contracontrol y no sobre sus condiciones, descalificando a quien lo emite, y por eso se trata como añadido. Que los cuatro funcionen como evitación en sentido técnico, mantenidos por evitar la corrección, es la parte que ninguna fuente del capítulo pone a prueba: describen los repertorios y, como mucho, la lectura de los propios autores, sin registrar la corrección y lo que el hablante o el grupo hizo después.

FIGURA 6.2

Control, contracontrol y los repertorios de evitación: una hipótesis



Los repertorios 1, 3 y 4 debilitan condiciones que Delprato (2002, p. 196) asocia al contracontrol eficaz (control conspicuo, conducta del controlador accesible, fuente identificada). El repertorio 2 es un añadido propuesto aquí.

Discontinuo: hipótesis propuesta aquí. El término "evitación del contracontrol" no es de Delprato.

Nota. Esquema original del autor, sin datos. Discontinuo: hipótesis. El término “evitación del contracontrol” no es de Delprato (2002). Los repertorios 1, 3 y 4 debilitan condiciones que él asocia al contracontrol eficaz; el repertorio 2 es un añadido propuesto aquí.

DATO	En grupos políticos de WhatsApp y Telegram observados en 2018, la crítica llevaba a la expulsión del autor, pocos miembros enviaban casi todo el contenido y los medios tradicionales se describían como no fiables (Evangelista & Bruno, 2019). La proporción de verificaciones de Lupa con WhatsApp subió de 2018 a 2024 (Agência Lupa, 2026). Guerin (1994) describe que los autoclíticos evitan consecuencias negativas de las declaraciones de actitud.
INTERPRETACIÓN	Los repertorios descritos son compatibles con lo que Delprato (2002) asocia al contracontrol ineficaz: control disfrazado, fuentes no identificadas y conducta del controlador inaccesible. Que reduzcan la probabilidad o la eficacia de la corrección es una lectura; que se mantengan solo por la aprobación de la audiencia es una lectura igualmente compatible.
HIPÓTESIS	Los repertorios de forma, fuente, lugar y personas se mantienen, al menos en parte, por evitar la corrección. Esto predice que la forma del discurso cambie después de las verificaciones, que el descrédito acompañe o preceda a la corrección y que la expulsión de disidentes o el desplazamiento de lugar la sucedan. Ninguna fuente del capítulo registra la corrección y lo que vino después. Aunque estos repertorios se presenten aquí en el contexto de la desinformación, la posición de los autores es que están presentes en cualesquiera interacciones sociales.

6.5 El caso brasileño: WhatsApp, elecciones y pandemia

El material brasileño del capítulo es escaso y procede de diseños diferentes. La Tabla 6.3 lo reúne antes de cualquier lectura.

TABLA 6.3

Material brasileño del capítulo: qué tiene cada fuente, qué registra y qué no permite decir

Fuente	Material	Qué registra	Qué no permite decir
Evangelista y Bruno (2019)	Observación participante en al menos tres grupos políticos de WhatsApp y dos de Telegram, hasta la elección de 2018 y las semanas anteriores; datos de Resende et al., citados por los autores	Expulsión de quien criticaba; pocos miembros enviaban casi todo el contenido; medios tradicionales descritos como controlados por comunistas e izquierdistas; autoría invisible	Pocos grupos, sin muestreo descrito: no mide prevalencia; los autores dicen no poder confirmar plenamente el uso eficaz de la microsegmentación
Bonow Soares et al. (2021)	Mensajes más compartidos de más de 500 grupos políticos públicos de WhatsApp (mar.-abr. de 2020): 802 con desinformación sobre la covid-19, según los autores	41% teorías de la conspiración, 39% distorsiones, 20% información fabricada; entre los tópicos (no excluyentes), 36% gobernadores y alcaldes con intereses ocultos contra Bolsonaro y 21% los medios	Solo grupos públicos, solo texto, cerca de un tercio de los mensajes compartidos al menos 20 veces por día; muchos grupos son de partidarios de Bolsonaro
Ricard y Medeiros (2020)	Comentario, sin datos propios; testimonios y CPMI	Parte de la desinformación atacaba a gobernadores, medios y especialistas en salud para deslegitimar a quienes apoyaban el aislamiento	No es un estudio empírico; la comisión aún no había divulgado evidencia de coordinación directa por parte del llamado Gabinete do Ódio (Gabinete del Odio) o de miembros del gobierno
Agência Lupa (2026)	286 verificaciones electorales (contenido viral sin autoría identificada), de 2018 a 2024	Facebook en el 76,5% (2018); WhatsApp en el 92,4% (2022) y el 100% (2024); 73,1% sobre urnas, recuento, TSE (Tribunal Superior Electoral) o resultados; 50% con instituciones públicas involucradas	Retrato del universo que Lupa monitorea (tipline en WhatsApp, programas de Meta); categorías no excluyentes
Recuero et al. (2025)	811 verificaciones de tres agencias, elecciones de 2018 y 2022	Cambio de plataforma leído como adaptación a espacios más cerrados; ambas campañas aparecen como productoras y víctimas	Estudio de verificaciones, no de grupos; lectura de los autores

Nota. Páginas de Evangelista y Bruno (2019), Bonow Soares et al. (2021), Ricard y Medeiros (2020), Agência Lupa (2026) y Recuero et al. (2025): las del PDF de las revistas y del informe. Los denominadores de Lupa, de 51 a 131 verificaciones por año, se leen en el gráfico de la p. 80.

En la tabla se leen dos cosas. La primera es lo que se repite entre fuentes de diseños diferentes: la descalificación de quien está fuera del grupo, de quien corrige o de quien apoya una política contraria (Evangelista y Bruno; Bonow Soares et al.; Ricard y Medeiros) y la autoría difícil de identificar (Evangelista y Bruno; Agência Lupa). En la pandemia, Bonow Soares et al. (2021) registran que la desinformación en los grupos monitoreados estuvo frecuentemente ligada a teorías de la conspiración, como supuestas tramas de izquierdistas, del STF (Supremo Tribunal Federal, la corte suprema de Brasil) y de los medios contra Bolsonaro. Dos codificadores clasificaron los 802 mensajes, con un tercero para desempatar: el 41% eran teorías de la conspiración, el 39% distorsiones y el 20% información fabricada, en categorías mutuamente excluyentes y con una confiabilidad entre codificadores (alfa de Krippendorff, medida de concordancia) de 0,601, que los autores califican de moderada a sustancial. Entre los tópicos, que podían superponerse, el 36% trataba de gobernadores y alcaldes con intereses ocultos contra Bolsonaro y el 21% de los medios. En la elección, la mitad de las verificaciones de Lupa involucra instituciones públicas, el Poder Judicial, las Fuerzas Armadas u órganos de seguridad, frecuentemente presentados como participantes en supuestas conspiraciones o irregularidades (Agência Lupa, 2026).

Hay una hipótesis en la que esos números encajan, y se propone aquí: si las instituciones que realizan el recuento y verifican la elección, la corte electoral, el Poder Judicial, la prensa, son las fuentes del contracontrol a las afirmaciones sobre eventos electorales, presentarlas de antemano como participantes en conspiraciones es un modo de debilitar el contracontrol que vendría de ellas, el repertorio de fuente de la sección 6.4. La lectura rival es que los contenidos solo tratan de lo que hay que tratar en una elección: el asunto de las urnas es el asunto de la elección, y las instituciones son su objeto. El material no separa las dos lecturas.

La segunda es el equilibrio. Casi todo el material brasileño del capítulo trata de grupos y campañas de un campo político, el de los partidarios de Bolsonaro, y eso se debe al diseño de las fuentes (Bonow Soares et al., 2021; Evangelista y Bruno, 2019), y no a una elección de campo, ya que el mismo método se aplica a todos. Recuero et al. (2025) registran que la literatura que citan señala más desinformación producida por partidarios de Bolsonaro en las dos elecciones, pero que ambas campañas aparecen como productoras y víctimas de desinformación. Evangelista y Bruno (2019) dicen que, en la muestra de Resende et al., no hay evidencia de que la coalición de izquierda haya usado las

mismas tácticas, aunque hay relatos de que la campaña de Haddad (Fernando Haddad, candidato del PT en 2018) también usó WhatsApp. Los dos registros no deciden si el fenómeno es asimétrico: dicen que la pregunta, para los repertorios de evitación, no se planteó con una muestra comparable. La hipótesis del capítulo es de mecanismo: los repertorios se definen por su relación con la corrección, y no por partido, y deberían aparecer dondequiera que haya un grupo cohesionado bajo el control de una audiencia que lo refuerza. La ausencia de datos no es ausencia de fenómeno, y el capítulo no afirma simetría; afirma que el dato no permite decirlo.

DATO	En grupos políticos de mensajería observados en 2018 y en grupos públicos de WhatsApp monitoreados en 2020, se describen expulsión de críticos, pocos productores, autoría invisible, desconfianza hacia los medios tradicionales y, en la pandemia, teorías de la conspiración que involucran a gobernadores, medios, izquierdistas y el STF (Evangelista & Bruno, 2019; Bonow Soares et al., 2021). La mitad de las verificaciones electorales de Lupa involucra instituciones públicas, frecuentemente presentadas como participantes en supuestas conspiraciones (Agência Lupa, 2026).
INTERPRETACIÓN	Lo que se describe es compatible con los cuatro repertorios de la sección 6.4, y la clasificación se propone aquí. Las fuentes tratan, en su mayoría, de un campo político, por diseño.
HIPÓTESIS	Los repertorios aparecerán, con la misma función, en grupos de otros campos cuando se les aplique la misma regla de muestreo (véase el protocolo del Capítulo 2, sección 2.9). No hay datos sobre esto en el capítulo.

6.6 Quién da “me gusta”, quién responde: un análisis exploratorio de dos videos políticos en Instagram

En octubre de 2026, durante la campaña de la segunda vuelta de la elección presidencial, se publicó en el perfil de Instagram de uno de los autores (@marcioborgesmoreira) una serie de videos cortos sobre política. El perfil tenía 11.550 seguidores y vivía de carruseles de divulgación científica en Análisis de la Conducta, vistos en general por algunos centenares de cuentas. Dos videos de la serie se salieron de la curva: “Eram só vovós com Bíblia?” (“¿Eran solo abuelitas con Biblia?”), sobre el 8 de enero de 2023 (el asalto a las sedes de los tres poderes en Brasilia), tuvo 1,29 millones de visualizaciones, y “Quanto custava o litro da gasolina em cada ano?” (“¿Cuánto costaba el litro de gasolina cada año?”), 735 mil. El autor esperaba ataques. El tema, la polarización de la

campaña y un público cien veces mayor que la lista de seguidores parecían reunir las condiciones de una audiencia negativa, en el sentido de Skinner retomado en la sección 10.1. Los ataques no llegaron: en los mensajes directos (Direct), nada; en los comentarios, por lo que se veía, casi solo aprobación. La impresión fue que los videos habían llegado a un millón de personas de la burbuja del autor, la del campo de Lula.

Esta sección examina esa impresión con los datos del propio perfil, y el caso interesa al capítulo por una razón precisa. Los estudios de la sección 6.2 miden lo que llega al feed y con quién se interactúa; ninguno mide las consecuencias que la audiencia dispone para lo que se dice ni lo que le ocurre a quien discrepa. En un perfil, quien publica tiene acceso a parte de eso: cuántas cuentas vieron, cuántas eran seguidoras, qué escribió cada persona, cuántos "me gusta" recibió cada comentario y qué respuestas provocó. El caso no pone a prueba la hipótesis del capítulo, porque es un perfil y un episodio, pero muestra lo que se gana al describir la burbuja por las consecuencias y dónde siguen faltando datos.

Hay una salvedad de origen, y viene antes de los números: los videos son de uno de los autores, que había tomado posición en la campaña. Eso no cambia las reglas de lectura del libro (dato, interpretación e hipótesis separados; el mismo criterio para todos los campos), pero el lector debe tenerlo en cuenta al juzgar la clasificación de los comentarios, que es interpretación.

El material y cómo se obtuvo

El material abarca las 74 publicaciones del perfil entre el 21 de septiembre y el 21 de octubre de 2026 y se leyó el 21 de octubre, entre las 20:00 y las 22:00 (hora de Brasilia). Procedió de tres fuentes. La Graph API de Instagram, la interfaz que Meta ofrece a las cuentas profesionales, proporcionó la fecha, el tipo, el texto, los "me gusta" y el número de comentarios de cada publicación y, en los dos videos, el texto de todos los comentarios y de las respuestas a ellos. La clave de acceso disponible no tenía permiso para leer métricas; por eso las visualizaciones, el alcance, la división entre seguidores y no seguidores, las veces compartido y guardado, el tiempo medio visto y el público por edad y género se leyeron en las páginas de cada publicación de Meta Business Suite, el panel de la propia Meta, y el perfil de los seguidores, en la página de público del mismo panel.

Dos medidas de la plataforma deben explicarse antes de usarse. El alcance es el número de cuentas diferentes que vieron la publicación; la visualización es cada vez que esta empieza a mostrarse, y por eso las visualizaciones superan al alcance (en el video del 8 de enero, 1,37 por cuenta). La plataforma también reparte visualizaciones y alcance entre seguidores y no seguidores del perfil,

y los dos repartos no cuadran: la página del video del 8 de enero atribuye 224.100 visualizaciones a seguidores, más de lo que las dos cuentas que lo publicaron (11.550 y 5.198 seguidores) producirían incluso con repeticiones, mientras que la página de la cuenta registra 5.202 seguidores alcanzados en todo el mes. Como Meta no explica la diferencia, el reparto se usa aquí solo por la dirección, que las dos medidas coinciden en mostrar.

Los comentarios llegaron incompletos, y la falta es en sí misma un dato. De los 2.529 que la plataforma contaba en el video del 8 de enero, la API devolvió 2.076 (1.901 de primer nivel y 175 respuestas), escritos por 1.778 cuentas; de los 962 del video de la gasolina, 876 (637 y 239), de 705 cuentas. Faltan 453 (18%) y 86 (9%). Ninguno de los devueltos estaba marcado como oculto, y la API no dice qué falta: comentarios borrados por quienes los escribieron, cuentas restringidas y comentarios retenidos por los filtros automáticos de la plataforma son explicaciones posibles, que el panel de la aplicación permitiría comprobar y que no se comprobaron. Los comentarios, con el nombre de usuario, quedaron solo en el servidor del autor; aquí no aparece ningún nombre, y los dichos hostiles se presentan parafraseados.

Cada comentario recibió tres códigos, según las reglas del Recuadro 6.1: la posición respecto al video, el campo político que declara y, si la hubiera, la hostilidad y su blanco. La lectura la hizo un solo codificador, el asistente de inteligencia artificial usado en la preparación de este libro, comentario por comentario (2.457 de ellos); otros 495, solo con mención a otra cuenta, solo con un sticker, solo con emojis de aplauso o risa o solo con el eslogan de un candidato, se clasificaron por regla. No hubo una segunda codificación independiente, y la concordancia entre codificadores no se midió; los números de la Tabla 6.5 valen con esa salvedad.

RECUADRO 6.1

Reglas de codificación de los comentarios

Posición respecto al video (una por comentario):

- **Adhesión al contenido:** está de acuerdo con lo que el video muestra o sostiene, elogia, agradece, añade un argumento o un hecho en la misma dirección o ironiza sobre la versión que el video cuestiona.
- **Manifestación de campaña:** voto, eslogan, hashtag o petición de voto, sin referencia al contenido del video.
- **Cuestionamiento:** discrepa, corrige, contraargumenta, relativiza o contrapone otro caso (“¿y el MST?”), cualquiera que sea el campo de quien cuestiona.
- **Ataque al autor:** insulto, descalificación personal o acusación de mala fe dirigidos a quien hizo el video.

- **Sin posición:** solo mención a otra cuenta, sticker, emoji ambiguo, pregunta neutra, relato de un precio local sin cuestionamiento, asunto ajeno.

Campo declarado: Lula e izquierda, Bolsonaro y derecha, o ninguno; solo cuando el comentario lo dice (voto, eslogan, apoyo o ataque nominal a un candidato, rótulo del otro campo, como "gado" [ganado] o "petralha" [apodo despectivo de los petistas]).

Hostilidad y blanco: insulto, rótulo despectivo de persona o grupo, deshumanización, deseo de daño o de castigo más allá del que decidió la Justicia, burla de un atributo personal (edad, apariencia). Blancos: participantes del 8 de enero; electores o militantes de derecha; políticos de derecha; Lula, el PT, la izquierda o sus electores; el autor; otro comentarista; otro. No cuentan los términos que describen actos ("vándalos", "golpistas", "criminales") ni los juicios sobre ideologías ("el bolsonarismo es una secta"). "Discurso de odio" no es un código: en el libro, es un rótulo de una instancia externa (Capítulo 5).

Ironía: solo se reconoce con un marcador (risa, comillas, "contiene ironía") o cuando la lectura literal contradice el propio comentario; sin eso, vale la lectura literal.

Grafía disfrazada: letra cambiada por número, @ o asterisco, o palabra partida por puntos o espacios ("c@gona", "L U L A"). Una regla automática encuentra los candidatos; la lectura decide.

Lo que muestran los números del perfil

El perfil tenía una rutina antes de la serie, y es frente a ella como se miden los dos videos. En los 45 carruseles de ciencia del período, la mediana fue de 464 visualizaciones y 202 cuentas alcanzadas, y solo el 3,7% de las visualizaciones vino de quienes no seguían el perfil: eran publicaciones vistas por la base. La serie política empezó la noche del 14 de octubre y sumó 24 Reels en siete días. Los 14 Reels de datos, todos en el formato "E quem estava na Presidência" ("Y quién estaba en la Presidencia": una serie anual de fuente oficial, con la foto de cada presidente), tuvieron una mediana de 3,3 mil visualizaciones, el 86,8% de ellas de no seguidores; los otros 8 Reels políticos, 2,4 mil y 76,8%. Los dos videos que se salieron de la curva sumaron 2,02 millones de visualizaciones, el 93% de todo lo que tuvieron las 74 publicaciones en el mes (Tabla 6.4 y Figuras 6.3 y 6.5).

TABLA 6.4

Los dos videos y la rutina del perfil: medianas por tipo de publicación, del 21/9 al 21/10/2026

Medida	Carruseles de ciencia (45)	Reels de datos (14)	Otros Reels políticos (8)	Video del 8 de enero	Video de la gasolina
Visualizaciones	464	3,3 mil	2,4 mil	1.293.276	735,5 mil

Medida	Carruseles de ciencia (45)	Reels de datos (14)	Otros Reels políticos (8)	Video del 8 de enero	Video de la gasolina
Cuentas alcanzadas	202	2,6 mil	2,0 mil	947.057	536,3 mil
Visualizaciones de no seguidores	3,7%	86,8%	76,8%	82,5%	93,9%
“Me gusta”	17	61	63	77.673	17.015
Comentarios	0	1	1,5	2.529	962
Veces compartido	1	12,5	14	23.232	14,7 mil
Veces guardado	6	9	9	8.864	2,1 mil
Tiempo medio visto	no se aplica	16 s	17 s	26 s (de 59 s)	25 s (de 49 s)
Hombres entre las cuentas alcanzadas	49,7%	79,7%	68,4%	54,0%	84,9%
Seguidores ganados	0	1,5	3	144	735

Nota. Medianas en las tres primeras columnas. Reels publicados del 14 al 21/10/2026; la columna de los carruseles abarca todo el mes, y la proporción de no seguidores tiene 44 publicaciones, porque la página de una de ellas no la mostraba. En las publicaciones sin publicación cruzada en Facebook, la plataforma da el alcance redondeado (“2,6 mil”). El video del 8 de enero salió en colaboración con otra cuenta, de 5.198 seguidores. Fuentes: Graph API de Instagram y Meta Business Suite, leídos el 21/10/2026; tabla completa, por publicación, en `verificacao/caso-instagram-2026/posts_metricas.tsv`.

El cambio de escala aparece entero en el alcance diario de la cuenta (Figura 6.4). Hasta el 13 de octubre, el perfil alcanzaba de 200 a 800 cuentas por día; el único salto del período fue un anuncio pagado de cinco días, del 30 de septiembre al 4 de octubre, para una publicación anterior a la serie (R$ 139,92, 23.502 visualizaciones). El 15 de octubre, primer día completo del video del 8 de enero, fueron 563.637 cuentas. Ninguno de los dos videos fue promocionado. El del 8 de enero superó las 500 mil visualizaciones unas 19 horas después de publicado y el millón en dos días y medio; el de la gasolina llegó a 500 mil en unas 39 horas y a 700 mil en cuatro días. En los 28 días del 23 de septiembre al 20 de octubre, la cuenta alcanzó 1,5 millones de cuentas, de las cuales 5.202 eran seguidoras (0,3%) y 1.512.615 no lo eran (99,7%); el 97% de las interacciones vino de no seguidores.

FIGURA 6.3

Los dos videos que se salieron de la curva y la serie de la que formaban parte

"¿Eran solo abuelitas con Biblia?"
14/10, 23:29, en colaboración



visualizaciones **1.293.276** | "me gusta" **77.673**
cuentas alcanzadas **947.057** | comentarios (leídos) **2.529 (2.076)**
visual. de no seguidores **82,5%** | veces compartido **23.232**
tiempo medio visto **26 s de 59 s** | veces guardado **8.864**

"¿Cuánto costaba el litro de gasolina?"
17/10, 11:00



visualizaciones **735,5 mil** | "me gusta" **17.015**
cuentas alcanzadas **536,3 mil** | comentarios (leídos) **962 (876)**
visual. de no seguidores **93,9%** | veces compartido **14,7 mil**
tiempo medio visto **25 s de 49 s** | veces guardado **2,1 mil**

La misma serie, en los mismos días: ocho Reels de datos y las visualizaciones de cada uno

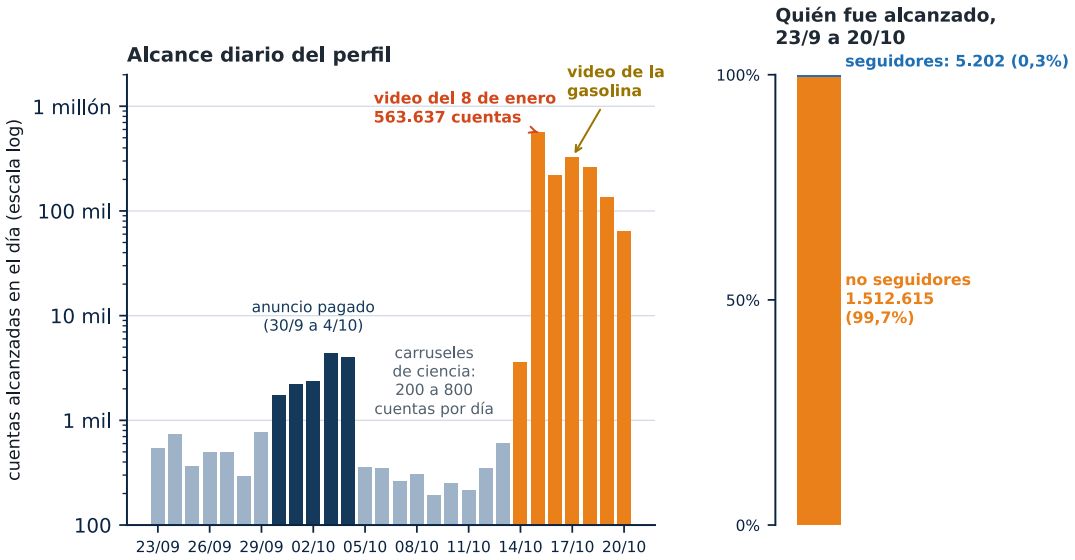


6,9 mil | 4,3 mil | 8,1 mil | 1,9 mil | 4,5 mil | 5,3 mil | 2,0 mil | 5,0 mil

Nota. Portadas de las publicaciones del autor, obtenidas mediante la Graph API. En la del 8 de enero, la mitad superior se desenfocó: muestra a una persona identificable, que no es figura pública, en un fotograma de video con la marca de un periódico. Fotos oficiales en las demás portadas: Ricardo Stuckert/PR, Roberto Stuckert Filho/PR, Beto Barata/PR y Alan Santos/PR (CC BY), acreditadas en los textos de las publicaciones. Números de Meta Business Suite y de la Graph API, leídos el 21/10/2026.

FIGURA 6.4

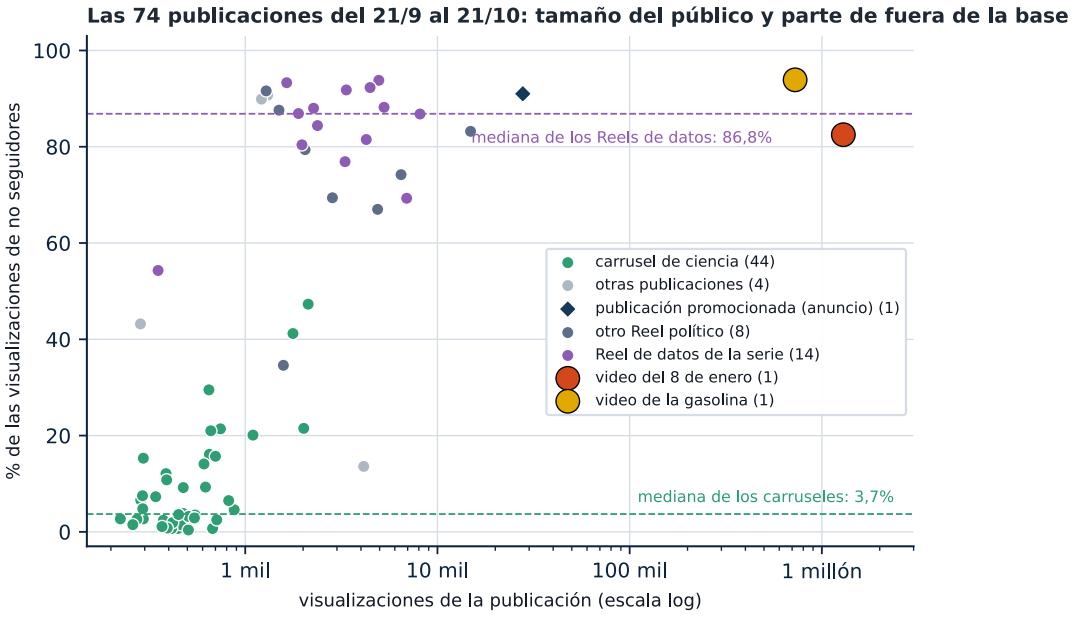
Alcance diario del perfil y proporción de seguidores y no seguidores, del 23/9 al 20/10/2026



Nota. Días contados en el huso horario del Pacífico, como en el panel de Meta: el video del 8 de enero, publicado a las 23:29 del 14/10 en Brasilia, cae al final del día 14/10 de ese huso. Fuente: Meta Business Suite, vista general de la cuenta de Instagram, leída el 21/10/2026.

FIGURA 6.5

Las 74 publicaciones del período: tamaño del público y proporción procedente de fuera de la base

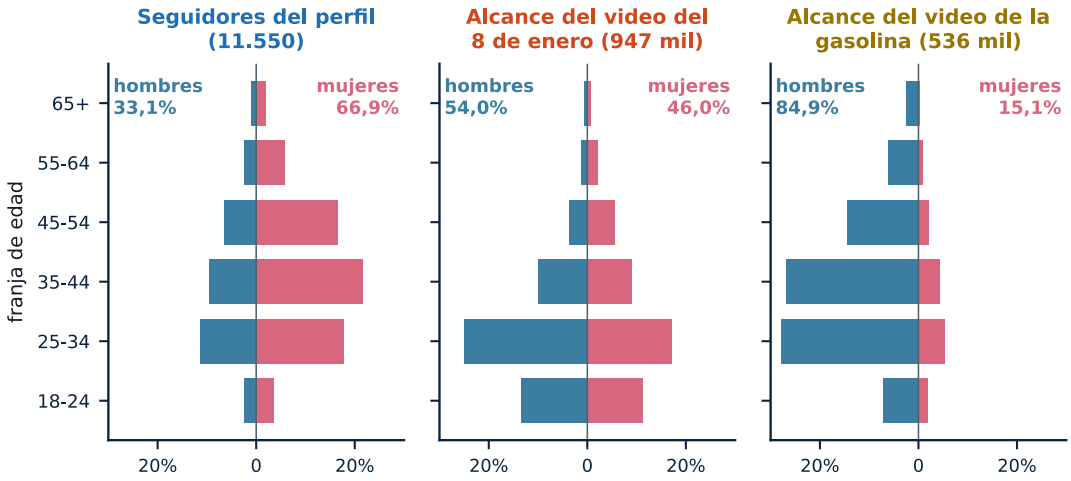


Nota. Cada punto es una publicación; la proporción de no seguidores es la de las visualizaciones, tal como la reparte la plataforma. “Otras publicaciones”: tres Reels institucionales (dos de libros del Instituto y uno de la Rádio iW4) y una imagen sobre el voto. Fuente: Meta Business Suite, páginas de cada publicación, leídas el 21/10/2026.

Quienes fueron alcanzados tampoco eran la base en otro sentido, el de la composición (Figura 6.6). Los seguidores del perfil son en un 66,9% mujeres, y el 83% de ellos tiene de 25 a 54 años. Entre las 947 mil cuentas alcanzadas por el video del 8 de enero, el 54% eran de hombres, y el 67% tenía de 18 a 34 años; en el video de la gasolina, el 84,9% eran de hombres. Incluso comparado con el público que alcanzaban los carruseles de ciencia, que tenía una mediana del 49,7% de hombres, el de los Reels de datos (79,7%) y el de la gasolina eran otro público.

FIGURA 6.6

Quién sigue el perfil y a quién alcanzaron los dos videos, por edad y género



Nota. Porcentajes del total de cada público. Seguidores: página de público de Meta Business Suite; alcance: página de cada publicación. La plataforma infiere la edad y el género a partir del registro de las cuentas. Leídos el 21/10/2026.

Otras dos medidas dicen qué hizo esa audiencia además de ver. Compartió y dio "me gusta" mucho, pero casi no se hizo seguidora: el video del 8 de enero dio al perfil 144 seguidores nuevos, uno por cada 6,6 mil cuentas alcanzadas, y el de la gasolina, 735; en el mes, hubo cerca de mil seguidores nuevos y 404 que dejaron de seguirlo. Y no le escribió al perfil: el panel registra 36 conversaciones iniciadas en el Direct en 28 días, 16 de ellas en los siete días de la serie, y 12 contactos nuevos, nada que acompañe un alcance de 1,5 millones de cuentas. El contenido de los mensajes no se leyó.

Lo que dijeron los comentarios

Los dos videos tuvieron comentarios de composición muy diferente (Tabla 6.5 y Figura 6.7). En el del 8 de enero, el 52,1% eran de adhesión al contenido, el 19,3% de campaña de Lula y el 26,0% sin posición; el cuestionamiento sumó el 2,2% (45 comentarios), la mitad de él de quienes declaraban el campo de Bolsonaro, y solo un comentario atacó al autor. Los comentarios que declaraban

el campo de Bolsonaro fueron 32 (1,5%), frente a 536 (25,8%) del campo de Lula. La adhesión tenía formas que se repetían: ironizar sobre la versión de las "abuelitas" ("pobrecitos", "solo había santos"), repetir la frase de la señora que muestra el video, señalar que los propios participantes grabaron las imágenes, comparar con protestas de la izquierda y pedir que no hubiera amnistía. La campaña llegaba en números, eslóganes y hashtags contra el adversario; una sola cuenta pegó el mismo mensaje de voto 30 veces.

En el video de la gasolina, un tercio de los comentarios (35,7%) cuestionó el video, y el cuestionamiento llegó desde tres lados. De quienes declaraban el campo de Lula llegaron 117 comentarios que decían que el video mentía porque, en el gobierno de Bolsonaro, el litro había llegado a 9, 10 o 12 reales: precios de pico, de una gasolinera y de una ciudad, frente a una media nacional anual corregida por la inflación, que la portada y el texto anunciaban como "en reales de hoy". Del campo de Bolsonaro llegaron 118, casi todos recordando la pandemia, la guerra en Ucrania y la rebaja de impuestos de 2022, que el video no mencionaba. Y 78, sin campo declarado, cuestionaron con los precios nominales de la época ("en 2006 estaba a 2,50"), muchas veces copiando el resumen de un buscador. Los comentarios del campo de Bolsonaro fueron el 13,9% del total, nueve veces la proporción del otro video.

TABLA 6.5

Los comentarios de los dos videos por categoría, y los "me gusta" que la audiencia dio a cada una

Categoría	8 de enero: comentarios	8 de enero: "me gusta"	Gasolina: comentarios	Gasolina: "me gusta"
Adhesión al contenido	1.081 (52,1%)	30.321 (91,0%)	202 (23,1%)	711 (24,3%)
Campaña de Lula	400 (19,3%)	820 (2,5%)	138 (15,8%)	414 (14,1%)
Sin posición	539 (26,0%)	2.062 (6,2%)	217 (24,8%)	1.008 (34,4%)
Cuestionamiento desde el campo de Lula	1 (0,0%)	0	117 (13,4%)	565 (19,3%)
Cuestionamiento sin campo declarado	22 (1,1%)	4 (0,0%)	78 (8,9%)	183 (6,2%)
Cuestionamiento desde el campo de Bolsonaro	22 (1,1%)	103 (0,3%)	118 (13,5%)	49 (1,7%)
Campaña de Bolsonaro	10 (0,5%)	3 (0,0%)	4 (0,5%)	0
Ataque al autor	1 (0,0%)	0	2 (0,2%)	0
Total	2.076	33.313	876	2.930
Con hostilidad, cualquier blanco	265 (12,8%)	941 (2,8%)	110 (12,6%)	168 (5,7%)

Nota. Categorías del Recuadro 6.1, con posición y campo combinados. Comentarios y respuestas que devolvió la Graph API (2.076 de 2.529 y 876 de 962); "me gusta" dados a esos comentarios hasta el 21/10/2026, por la noche. En la gasolina, la categoría "sin posición" concentra "me gusta" por un sticker que recibió 849. Codificación de un solo codificador, sin medida de concordancia. Fuente: `verificacao/caso-instagram-2026/comentarios_categorias.tsv`.

La hostilidad existió en los dos videos, en proporción parecida (12,8% y 12,6% de los comentarios), y casi toda apuntó hacia fuera. En el del 8 de enero, los blancos fueron los participantes de los actos (117 comentarios: insultos a la señora del video, burlas sobre la edad, peticiones de una pena mayor que la decidida, algunos deseos de muerte), políticos de derecha (75), electores de derecha (31, con "gado" y "bolsominions"), otros comentaristas (24), Lula, el PT o la izquierda (5) y el autor (1). En el de la gasolina, otros comentaristas (39, casi todos en respuestas), políticos de derecha (32), Lula, el PT o la izquierda (23), electores de derecha (13) y el autor (3). Sumados los dos videos, cuatro de 2.952 comentarios (0,14%) fueron hostiles hacia quien los publicó; uno de ellos, de alguien que se declaraba del campo de Lula, acusaba al video de mentir por mostrar el precio del gobierno de Bolsonaro más bajo de lo que él recordaba. Quien llegó con hostilidad la dirigió, en general, al grupo que el video criticaba, y no a quien lo publicó.

Quién refuerza qué: "me gusta" y respuestas

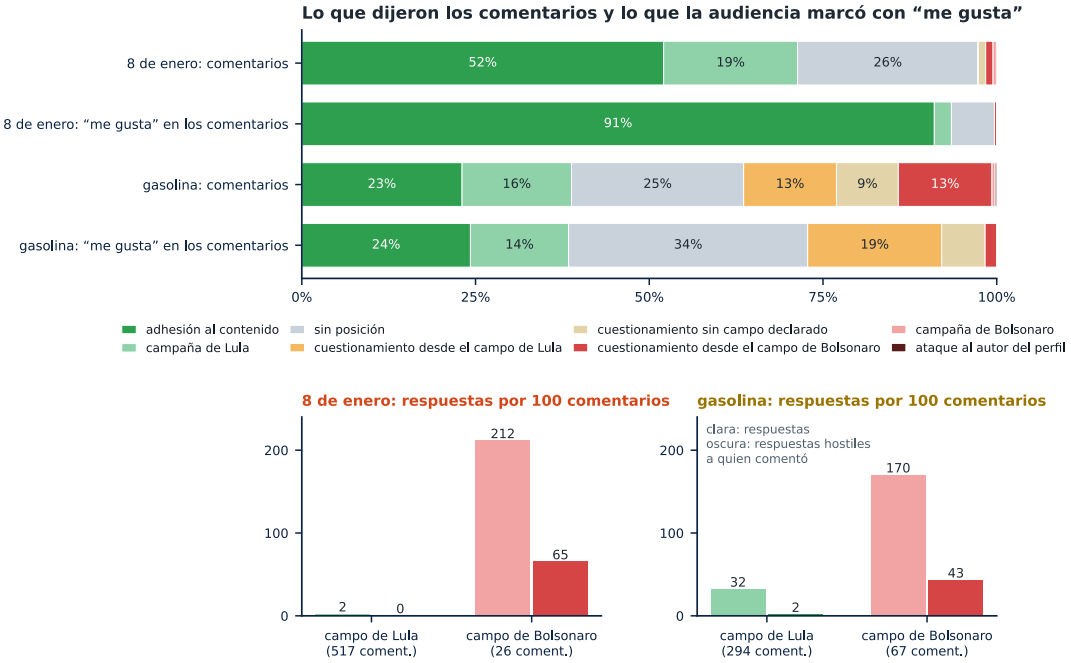
La audiencia no solo comenta: da "me gusta" a comentarios y les responde, y esas son las consecuencias que dispone, visibles y contadas, para lo que cada persona escribió. En el video del 8 de enero, los comentarios de adhesión recibieron el 91% de todos los "me gusta" dados a comentarios. Los cuatro con más "me gusta", todos de adhesión, tenían 12.623, 8.608, 3.999 y 1.266 (el que más tenía decía que "lo mejor es que ellos mismos lo grabaron"); como la plataforma ordena los comentarios por relevancia, eso era lo que se veía primero al abrirlos, y es probable que mi impresión de "casi solo aprobación" viniera de ese escaparate. En el video de la gasolina, los dos grupos de cuestionamiento partidario tenían el mismo tamaño, 117 y 118 comentarios, pero los del campo de Lula recibieron 565 "me gusta" y los del campo de Bolsonaro, 49. La audiencia reforzó el cuestionamiento por el campo de quien cuestionaba, y no por la calidad del argumento: recordar la pandemia no es menos pertinente que recordar un precio de pico. La corrección dentro del propio campo también fue reforzada: un comentario que decía que en el gobierno de Dilma la gasolina nunca había llegado a 4 reales tuvo 169 "me gusta", la respuesta que explicaba la corrección por la inflación tuvo 284, y quien había comentado respondió "entendí".

Las respuestas muestran el otro lado de la misma contingencia (Figura 6.7). En el video del 8 de enero, los 517 comentarios de primer nivel del campo de Lula recibieron, en conjunto, 12 respuestas, una de ellas hostil; los 26 del campo de Bolsonaro recibieron 55, de las cuales 17 eran hostiles hacia quien había comentado. Por cada cien comentarios, son 2 respuestas y menos de una hostil en un campo, 212 respuestas y 65 hostiles en el otro. En el video de la gasolina, donde la discusión fue mayor para todos, la diferencia persiste: 32 respuestas y 2 hostiles por cada cien comentarios del campo de Lula, 170 y 43 por cada cien del campo de Bolsonaro.

Dos hilos muestran cómo ocurre esto. En el video del 8 de enero, una comentarista escribió que los participantes "aprendieron del MST" (el Movimento dos Trabalhadores Rurais Sem Terra, movimiento de trabajadores rurales sin tierra), con dos emojis de risa. El comentario tuvo 94 "me gusta" y 47 respuestas: parte argumentaba (el MST produce alimentos, donó comida en la pandemia), y 15 eran hostiles hacia ella, entre ellas llamarla ignorante y mandarla callar, decir que quien debía opinar era el marido, comentar el color de su piel y la profesión del marido y, en una de ellas, una advertencia velada sobre las fotos de las hijas que ella publicaba. En el video de la gasolina, un comentarista escribió que en su ciudad el litro costaba 5 reales en el gobierno de Bolsonaro y 7,80 ahora. Recibió 35 respuestas, 18 de ellas hostiles, casi todas de burla; la que tuvo más "me gusta", 77, preguntaba si había milicia en las gasolineras de su ciudad. En un tercer episodio, el comentario "Flávio Bolsonaro es nuestro futuro presi..diario" (659 "me gusta") fue leído por algunos como apoyo a Flávio y recibió respuestas como "el tuyo, esa maldición", hasta que alguien avisó que no lo habían leído bien: la audiencia respondía a lo que parecía disidencia antes de leerla entera.

FIGURA 6.7

Lo que dijeron los comentarios, a qué dio "me gusta" la audiencia y qué les ocurrió a quienes discreparon



Nota. Comentarios y respuestas devueltos por la Graph API; "me gusta" hasta el 21/10/2026, por la noche. Abajo, solo los comentarios de primer nivel con campo declarado y las respuestas que recibieron; respuesta hostil es la que tiene hostilidad dirigida a quien comentó (Recuadro 6.1). Codificación de un solo codificador. Fuente: `verificacao/caso-instagram-2026/comentarios_categorias.tsv` y `respostas_por_campo.tsv`.

Forma, fuente y personas: los repertorios de la sección 6.4 en la caja de comentarios

Los cuatro repertorios de la sección 6.4 se describieron, en el material brasileño del capítulo, en grupos de mensajería de un solo campo político. Tres de ellos aparecen en la caja de comentarios de estos videos, en una audiencia alineada con el campo opuesto, y la descripción que sigue es de topografía: dice qué forma tenían y qué consecuencias siguieron, y no que el repertorio se mantenga por evitar la corrección.

Forma. Veintiocho comentarios (0,9%) disfrazaban la grafía: "c@gona", "vaga@bund0s", "m4t4e", "ban.did0 bom e ba.ndid0 m0.rt0" ("delincuente bueno es delincuente muerto"), "presi..diário", "L U L A". El disfraz recaía sobre insultos, deseos de muerte y nombres de candidatos, y un comentario declaraba abiertamente la función: eviten escribir el nombre de Lula, porque el algoritmo nos está bloqueando. Es una forma que reduce la probabilidad de que la moderación automática de la plataforma, un control que viene de fuera de la comunidad, alcance lo que se escribe, y el comentario que pedía el dis-

fraz muestra que al menos una persona describía así la contingencia. Que la moderación de hecho bloqueara esos comentarios, ningún dato del caso lo muestra.

Fuente. En la gasolina, cuestionamientos de los dos campos llamaron mentira al video, y las respuestas a quien discrepaba llamaron fake news a la discrepancia ("esas fake news no cuelan aquí"). La descalificación de la fuente del contracontrol, que la sección 6.4 describió en los grupos de 2018 a propósito de los medios tradicionales, aparece aquí a la escala del comentario, dirigida a quien aporta el dato contrario.

Personas. En la caja de comentarios no hay un administrador que expulse a quien critica, como en los grupos observados por Evangelista y Bruno (2019), pero existe el equivalente que la audiencia puede disponer: responder en masa, con burla e insulto, al comentario que discrepa. Los números de la sección anterior son esa consecuencia, contada.

Lugar. No hubo desplazamiento a un espacio cerrado; hubo el movimiento contrario. Comentarios de campaña de Lula llevaban hashtags del campo adversario ("#bolsonaro2022", "#fláviobolsonaropresidente"), lo que pone el mensaje en las búsquedas de quienes siguen esas etiquetas. Es una lectura de la función, y el caso no mide si alguien del otro campo llegó al video a través de ellas.

¿Era la burbuja del perfil?

La pregunta de la apertura pide tres respuestas, porque "burbuja" abarca tres cosas diferentes. Si la burbuja es la lista de seguidores, los videos no se quedaron en ella: el 99,7% de las cuentas alcanzadas no seguía el perfil, y el público alcanzado era más joven y más masculino que la base. Si la burbuja es quien vio, el caso no sabe decirlo: el panel no informa la posición política de nadie, y los 2.076 comentarios leídos del video del 8 de enero los escribió el 0,2% de las cuentas que alcanzó. Si la burbuja es la subcomunidad verbal de la sección 6.1, los oyentes que reaccionan a lo que se dice y disponen consecuencias para ello, la respuesta es sí, y tiene números: en el video del 8 de enero, el 1,5% de los comentarios declaraba el campo opuesto al del video, los "me gusta" fueron a la adhesión y los pocos que discreparon recibieron respuestas en una proporción cien veces mayor que los demás, de una cuarta parte a un tercio de ellas hostiles. Guerin (1992) escribe que los cambios en el uso del lenguaje pueden estar bajo el control de subcomunidades verbales más pequeñas que la comunidad verbal amplia; esta fue una, montada en dos días, que no existía como grupo antes del video y no se hizo seguidora después de él.

Cómo una audiencia de un millón de cuentas de fuera de la base puede ser tan homogénea en lo que dice es la pregunta que deja el caso, y tiene cuatro respuestas rivales. La primera es la distribución: el ordenamiento de la plataforma decide cuántos ojos y cuáles (sección 8.3), y, si usa las reacciones de quienes ya vieron para elegir a quién mostrarlo después, la audiencia se va seleccionando por sus propias reacciones; ninguna fuente del libro muestra qué usa el ordenamiento como entrada, y el caso tampoco. La segunda es la audiencia negativa: quienes eran del otro campo vieron y no comentaron, porque la caja de comentarios castigaba a quien discrepaba. Skinner (1957) afirma que el efecto de una audiencia negativa solo se detecta en combinación con una audiencia positiva, como reducción de la fuerza de la conducta apropiada a esta, y que, mientras las audiencias permanecen separadas, el dilema no aparece; los hilos de respuestas muestran la consecuencia que esperaba a quien discrepara, pero no cuántos dejaron de comentar por eso. La tercera es el filtrado: los 453 y 86 comentarios que la API no devolvió, los filtros de comentarios ofensivos y las carpetas de mensajes ocultos del Direct pueden haber retenido justamente los ataques, y eso no se comprobó. La cuarta es el contenido: imágenes de las cámaras y dichos literales, en el video del 8 de enero, dejan menos que cuestionar que una serie de precios, y el video de la gasolina, con datos que cada uno confronta con su propia memoria, tuvo nueve veces más comentarios del campo adversario. Las cuatro pueden valer a la vez, y el caso no las separa.

El caso dialoga con los estudios de la sección 6.2 en un punto preciso. Nyhan et al. (2023) muestran que el contenido de fuentes de la misma inclinación es la mayor parte de lo que ven los usuarios de Facebook, y que quienes tuvieron esa exposición reducida interactuaron más, por exposición, con el contenido afín que aún veían. Aquí la exposición no vino de la red de quién sigue a quién, y aun así la interacción fue homofílica, en el sentido de Cinelli et al. (2021): quienes reaccionaron tenían, en abrumadora mayoría, la inclinación del contenido. Que la homofilia de las interacciones dependa menos de la exposición que de las consecuencias que dispone la audiencia es la hipótesis del capítulo, y el caso es compatible con ella; la selección por la distribución, la primera rival de arriba, es igualmente compatible.

Las cinco preguntas del Recuadro 6.3 dicen hasta dónde llega el caso. La corrección existió y es identificable: los comentarios del campo adversario y los cuestionamientos sin campo corregían o cuestionaban el video, y la respuesta en masa a ellos es la consecuencia que la audiencia dispuso para quien corregía. El discurso posterior no se registró: el caso es un retrato de una noche, sin el antes y el después de quienes recibieron respuestas, y no dice si esas personas volvieron a comentar. La audiencia no es la misma de otro momento, por-

que la plataforma la montó para estos videos. El reforzamiento diferencial se midió, en los "me gusta" y en las respuestas. Y la rival de la aprobación sin evitación sigue en pie: los "me gusta" y las respuestas bastan para describir lo que se vio, sin suponer que alguien evitó la corrección. Llamar evitación del contracontrol a lo que ocurrió en la caja de comentarios, tal como el capítulo define el término, exigiría la segunda pregunta, que el caso no responde.

DATO	Dos videos de un perfil de 11.550 seguidores, publicados en la campaña de 2026, sumaron 2,02 millones de visualizaciones; en el mes, el 99,7% de las cuentas alcanzadas por el perfil no lo seguía, y el público alcanzado era más joven y más masculino que la base. En los 2.952 comentarios leídos, la declaración del campo adversario al del video fue del 1,5% (8 de enero) y del 13,9% (gasolina). La adhesión recibió el 91% de los "me gusta" dados a comentarios en el primer video; en el segundo, el cuestionamiento desde el campo de Lula recibió 565 "me gusta" y el del campo de Bolsonaro, del mismo tamaño, 49. Los comentarios del campo adversario recibieron de 170 a 212 respuestas por cada cien, y de 43 a 65 hostiles; los del campo del video, de 2 a 32 y de 0 a 2. Cuatro comentarios fueron hostiles hacia el autor. La API no devolvió el 18% y el 9% de los comentarios contados.
INTERPRETACIÓN	Los comentaristas funcionaron como una subcomunidad verbal montada para cada video, que reforzó la adhesión y respondió a la discrepancia con consecuencias aversivas. La ausencia de ataques al autor es compatible con cuatro lecturas, la distribución selectiva, la audiencia negativa, el filtrado por la plataforma y la forma del contenido, que el caso no separa.
HIPÓTESIS	Si la respuesta en masa a la discrepancia funciona como castigo, quien recibió respuestas comentará menos después, y los videos sobre temas en los que cada lado confronta el dato con su propia experiencia (precios) atraerán más comentarios del otro campo que los videos con imágenes de los hechos. Ponerla a prueba requiere el historial de comentarios de las mismas cuentas antes y después, y la comparación con perfiles de otros campos bajo la misma regla de codificación.

Para quien quiera repetir el análisis en un perfil, de cualquier campo, el Recuadro 6.2 reúne los pasos que el caso mostró necesarios.

RECUADRO 6.2

Cómo medir la burbuja de un perfil por las consecuencias

1. Defina el período y una línea de base: las publicaciones habituales del perfil, con alcance y proporción de no seguidores, antes del episodio.
2. Lea, por publicación y para la cuenta, el alcance de seguidores y de no seguidores y el público por edad y género; anote la hora de la lectura, porque los números cambian.
3. Recoja todos los comentarios y respuestas mediante la API y registre cuántos cuenta la plataforma y cuántos devuelve; antes de concluir que no hubo ataques, compruebe en la aplicación los comentarios ocultos, el filtro de palabras ocultas y las solicitudes de mensajes ocultas.
4. Codifique posición, campo declarado y hostilidad con reglas escritas (Recuadro 6.1), y codifique aparte, con un segundo codificador, una muestra sorteada, para medir la concordancia.
5. Mida las consecuencias: "me gusta" por categoría de comentario, respuestas y respuestas hostiles por campo de quien comentó, y los hilos en los que se respondió a la discrepancia.
6. Antes de llamar evitación a un repertorio, pase por las cinco preguntas del Recuadro 6.3; sin el discurso posterior de quienes recibieron respuestas, lo que se tiene es una descripción de consecuencias, y no de evitación.

6.7 Análisis semánticos en redes sociales

Uno de los libros del autor (Moreira et al., 2022a), que reúne también estudios con cómics (historietas) y con autoritarismo (Capítulo 4, sección 4.7), presenta un estudio de análisis semántico de publicaciones espontáneas, el Estudio 4. La pregunta era si, a partir de textos de redes sociales, sería posible describir clases de equivalencia ligadas al movimiento *Black Lives Matter* y, con ellas, la visión social del movimiento, medida por la influencia semántica de las palabras usadas. La recolección fue automática, mediante scripts en R que consultaron la API pública de Twitter durante 20 días, del 22 de abril al 11 de mayo de 2021, filtrada por hashtags del movimiento. Antes del análisis, se retiraron los nombres de usuario (de quien publicó y de quien fue mencionado) y todas las palabras que no fueran nombre propio, adjetivo, sustantivo o verbo. Las que quedaron se contaron y clasificaron por clase gramatical y por influencia semántica: positiva, negativa o "no se aplica"; el capítulo no describe el criterio ni quién atribuyó la influencia. Los resultados se presentan en gráficos de frecuencia, sin números en el texto: predominaron los sustantivos, y los adjetivos fueron pocos; la mayor parte de las palabras quedó en "no se aplica", y, entre las positivas y las negativas, las negativas fueron más numerosas, incluso entre los adjetivos; las palabras más frecuentes se relacionaban con el jui-

cio del caso George Floyd (“caso”, “jurado”, “decisión”, “justicia”). Como la recolección tuvo lugar en la semana del veredicto de ese juicio, los autores concluyen que no se puede afirmar que la visión social del movimiento sea negativa, aunque haya más palabras de influencia negativa que positiva, y sugieren repetir la recolección en el momento de la muerte de Floyd. Registran dificultades, como publicaciones incompletas y la separación manual de las palabras.

Los autores proponen, a partir de ahí, una expansión conceptual: clases de equivalencia “compartidas (o comunes)”, encontradas en muchos individuos, que pasan de fenómeno individual a fenómeno de alta frecuencia y se aproximan al concepto de representación social. Según los autores, recoger publicaciones espontáneas, sin que quien publica supiera que serían analizadas, excluye la variable de deseabilidad social, a diferencia de los cuestionarios de los estudios 1 y 2 del mismo libro.

Para este capítulo, la propuesta importa por una razón y exige una salvedad. La razón es que, si la burbuja es una subcomunidad verbal, la hipótesis natural es que las clases compartidas dentro de ella sean más homogéneas que las de la comunidad más amplia, y el método del estudio, que mide la influencia semántica de las palabras en un corpus de publicaciones espontáneas, es un modo de ponerla a prueba: bastaría separar el mismo corpus por comunidad (las clases de equivalencia están en el Capítulo 4, sección 4.2). La salvedad tiene dos partes. El estudio no separó comunidades, porque se retiraron los nombres de usuario y la recolección fue por hashtag, y por tanto no dice nada sobre burbujas. Y la espontaneidad de la publicación no elimina la audiencia: quien publica en un feed habla a sus seguidores, y la audiencia controla lo que se publica, como en el ciclo del Capítulo 9. Lo que el estudio excluye, la deseabilidad social ante un investigador, es una audiencia, y no la ausencia de audiencia.

6.8 Interpretaciones rivales y límites

La lectura del capítulo, de que la burbuja es una subcomunidad verbal y de que parte de lo que se mantiene en ella puede ser evitación del contracontrol, tiene rivales. El recuadro de abajo las reúne, una por una, y los límites del capítulo vienen después de él.

Interpretaciones rivales para el mismo conjunto de datos

Selección, y no control. La homofilia que miden Cinelli et al. (2021) puede reflejar quién elige con quién hablar, y no lo que la comunidad hace con lo que se dice: las personas que ya están de acuerdo se acercan, y la burbuja sería consecuencia del acuerdo, y no su causa. Los estudios del capítulo no separan las dos direcciones.

Aprobación sin evitación. Los repertorios de la sección 6.4 pueden mantenerse solo por el reforzamiento positivo de la audiencia, "me gusta", reenvíos, pertenencia (Capítulo 9), sin que la corrección sea aversiva ni que el repertorio la evite. La aprobación basta para explicarlos. Distinguir las dos contingencias exige mostrar que el repertorio cambia después de la corrección, y ninguna fuente del capítulo lo registra.

Desconfianza como discriminación, y no como evitación. Quien descalifica a un corrector puede estar bajo el control de su historia con esa fuente. Freeze et al. (2021) mostraron, en un experimento en línea, que las advertencias inválidas de desinformación reducen la credibilidad de noticias auténticas y perjudican la memoria política. Participaron 434 personas de una plataforma de crowdsourcing (MTurk) en los Estados Unidos, que vieron un video de discursos de diputados en C-SPAN, respondieron a preguntas sin relación y leyeron un artículo periodístico fabricado sobre el video, unas veces vago, otras con hechos correctos, otras con parte de los hechos alterados. Una parte de ellos, sorteada, recibió a continuación la advertencia de que "parte de la información presentada en el artículo de noticias que usted leyó era inexacta", advertencia inválida cuando el artículo era exacto; después todos hicieron una prueba de reconocimiento de 20 hechos del video y evaluaron la credibilidad del video y del artículo. La advertencia hizo que el artículo pareciera 0,59 puntos menos creíble en una escala de 5, en todas las condiciones, incluso cuando era exacto, y el efecto sobre la memoria fue significativo en una de tres pruebas. Si las correcciones se equivocan o tienen costos, desconfiar de ellas no es, por sí mismo, un repertorio de protección, y el análisis necesitaría la tasa de acierto de la fuente descalificada, que el capítulo no tiene.

Razonamiento motivado. Bail et al. (2018) proponen que la exposición a mensajes que entran en conflicto con las actitudes provoca contraargumentación y razonamiento motivado. Es una explicación de los autores, en términos cognitivos, y ellos dicen que no identificaron el mecanismo. La lectura operante, de que lo que declara el participante está controlado por la audiencia y la historia, es una hipótesis que el experimento no pone a prueba.

Migración por uso, y no por moderación. Lupa lee el cambio de plataforma como una migración hacia ambientes semiprivados y, al mismo tiempo, advierte que sus resultados retratan el universo que ella monitorea (Agência Lupa, 2026). Evangelista y Bruno (2019) señalan que el *zero-rating* favorece la popularidad de

WhatsApp. La mayor presencia de WhatsApp puede reflejar el crecimiento del uso de la plataforma y del canal de recepción de la agencia, y no una huida de la moderación.

Muestras de un solo lado. El material brasileño abarca, en su mayor parte, grupos de un campo (Bonow Soares et al., 2021). Lo que se observa puede ser una propiedad de cualquier grupo cohesionado y no de este, y el capítulo no tiene cómo decidirlo.

Cuatro límites valen para todo el capítulo. Primero, ninguna fuente registra la corrección y lo que el hablante hizo después de ella, de modo que la evitación del contracontrol sigue siendo una hipótesis en todos los casos. Segundo, la exposición, la interacción y la participación no son consecuencias dispuestas por la audiencia, y los datos sobre cámaras de eco miden las primeras. Tercero, el material brasileño procede de una plataforma principal, WhatsApp, entre 2018 y 2022, de la observación de pocos grupos, de grupos públicos monitoreados y de un universo seleccionado por una agencia; generalizar a partir de ahí es extrapolación. Cuarto, el término "evitación del contracontrol", el agrupamiento en cuatro repertorios y las dos figuras se proponen aquí, y no proceden de una fuente.

Para quien quiera usar el capítulo en el análisis de discursos (Capítulos 3, 5 y 7), la consecuencia es práctica. Afirmar que un discurso está protegido de la corrección exige el registro de la corrección y el del discurso posterior, y una verificación, por sí sola, no aporta ese segundo término. El recuadro de abajo reúne qué preguntar antes de llamar evitación a un repertorio.

RECUADRO 6.3

Cinco preguntas antes de llamar evitación del contracontrol a un repertorio

1. ¿Cuál fue la *corrección*: quién la hizo, cuándo, con qué rótulo, y llegó al hablante o a la comunidad?
2. ¿El discurso *posterior* cambió de forma, de fuente, de lugar o de autoría respecto al anterior, y en cuánto tiempo?
3. ¿La *audiencia* es la misma, con la misma composición, antes y después?
4. ¿La audiencia *reforzó* el cambio (reacciones, reenvíos, permanencia) más que la forma anterior?
5. ¿Qué *rival* sigue en pie: aprobación sin evitación, discriminación de la fuente, cambio de audiencia o de plataforma?

Síntesis

La burbuja, vista como subcomunidad verbal, es un conjunto de oyentes con una historia de reforzamiento común. Los datos sobre cámaras de eco miden otra cosa: lo que llega al feed y con quién se interactúa. En tres estudios, la homofilia es grande en temas específicos, el feed entero es menos homogéneo de lo que supone el temor corriente, y ni la reducción de la exposición afín ni el aumento de la exposición al campo opuesto desplazaron las actitudes en la dirección esperada (sección 6.2). Ninguno de ellos mide quién refuerza lo que dice el hablante ni lo que de la corrección llega hasta él, de modo que la pregunta de la burbuja como comunidad verbal sigue abierta.

Guerin (1992) da la condición bajo la cual lo dicho contrafáctico se mantiene: un grupo pequeño y cohesionado, en el que pertenecer vale más que relatar con exactitud. Fuera de ella, la presión de conformidad produce contracontrol (6.4). La propuesta del capítulo es que hay repertorios del hablante y de la comunidad que reducen la probabilidad o la eficacia de ese contracontrol, la evitación del contracontrol, término propuesto aquí, de cuatro tipos: forma, fuente, lugar y personas. Tres debilitan condiciones que Delprato (2002) asocia al contracontrol eficaz; el cuarto es un añadido propuesto aquí (6.4). El material brasileño describe cada uno de los cuatro, con límites de muestra, de plataforma y de campo político que la sección 6.5 enumera. Lo que no aporta es la corrección y lo que el hablante hizo después; por eso la evitación sigue siendo una hipótesis, y la aprobación de la audiencia, sin evitación, es la rival más fuerte (6.8).

Los Capítulos 3 y 5 toman de este capítulo la idea de que la forma del discurso (actitud, autoclítico, mando disfrazado) puede mantenerse por reducir la corrección, y que eso se pone a prueba con discursos fechados, antes y después de las verificaciones; el Capítulo 8, que lo que la plataforma entrega al feed decide cuán permeable es la frontera de la Figura 6.1; el Capítulo 12, la pregunta de qué intervenciones actúan sobre las contingencias de la comunidad, y no solo sobre la exposición.

Referencias

Agência Lupa. (2026, 27 de julio). *Anatomia da desinformação eleitoral* [Informe]. 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), Artículo 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* (1.ª 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.

instituto WALDEN4 | ACADEMIA WALDEN4



LAB WALDEN4

Rehaga los experimentos clásicos en el navegador

Estudios clásicos de la psicología reconstruidos para funcionar en la computadora o en el teléfono móvil. Usted participa, y sus datos aparecen junto a los resultados originales.

8 experimentos para hacer

Acceda, gratis y sin registro

walden4.com.br/experimentos

CAPÍTULO

7

El voto como conducta verbal

Un candidato puede hablar con los electores hasta la víspera de la elección, pero tiene prohibido entrar con ellos en la cabina de votación. La observación es de Lloyd (1994), que la hace para ilustrar otro problema, el de que el agente de cambio puede tener acceso solo a la conducta verbal y nunca a la conducta que le interesa. Sirve de apertura aquí porque nombra la asimetría que organiza el capítulo: lo que el elector dice tiene audiencia, y el voto, por ser secreto (Visser, 1996), no la tiene en el momento en que ocurre.

Los Capítulos 3 y 5 analizaron lo que dice el político. Este capítulo se vuelve hacia quien lo escucha y vota. La pregunta es por qué votar puede tratarse como conducta verbal, es decir, como conducta mantenida por consecuencias que median otras personas, y qué cambia eso en la lectura de la relación entre lo que el elector dice y lo que hace en la urna. Los instrumentos son los del Capítulo 3, el operante verbal y la audiencia, y los del Capítulo 4, la actitud como conducta verbal evaluativa y la transferencia de función entre estímulos de una misma clase. Falta uno, que entra en la sección 7.4, en el punto en que el análisis lo necesita: la correspondencia entre decir y hacer.

Una advertencia sobre las fuentes. Ninguno de los estudios usados aquí mide el voto de electores brasileños. Visser (1996) analiza una encuesta electoral holandesa; Bond et al. (2012), una elección legislativa de Estados Unidos; los estudios de actitud y conducta son, en su mayor parte, de estudiantes estadounidenses, entre las décadas de 1930 y 1970. Los estudios brasileños disponibles, los del Capítulo 4, tratan de actitudes ante presidentes, y no de votos. El capítulo es, por lo tanto, una propuesta de análisis, y la extensión de los hallazgos a Brasil es una hipótesis, no un relato de lo que se sabe del voto brasileño.

7.1 El voto según la definición de conducta verbal

Para Skinner (1957), la conducta verbal es la conducta reforzada por la mediación de otras personas (el oyente), mientras que el conjunto de las conductas del hablante y del oyente es el "episodio verbal total", en el que nada "emerge" en la unidad social más allá de la conducta combinada de dos o más individuos. La definición no especifica forma, modo ni medio: cualquier movimiento capaz de afectar a otro organismo puede ser verbal. Señalar es verbal, porque solo es eficaz cuando altera la conducta de alguien.

La aplicación al voto se hace aquí, y conviene decir dónde es frágil. Marcar un voto es un movimiento que tiene efecto sobre un medio, la papeleta o la pantalla, y por medio de él sobre el resultado de la elección, sobre quién gobierna y sobre lo que la comunidad del elector hace a continuación. Según el criterio de Skinner, es verbal en la medida en que la consecuencia que lo mantiene está mediada por personas. La fragilidad está en que el oyente de esta conducta no es una persona presente que refuerza el acto, como lo es el oyente de una petición; es un sistema, y el reforzamiento del acto, en el sentido en que lo definimos para la conducta verbal, no ocurre en el momento en que se emite. El capítulo no resuelve la clasificación por decreto. Adopta una afirmación más modesta: el voto es una conducta cuyo mantenimiento es social, y por eso le corresponde el análisis de relaciones de control mediadas, porque lo que lo mantiene es lo que la comunidad del elector hace con él y con lo que el elector dice de él.

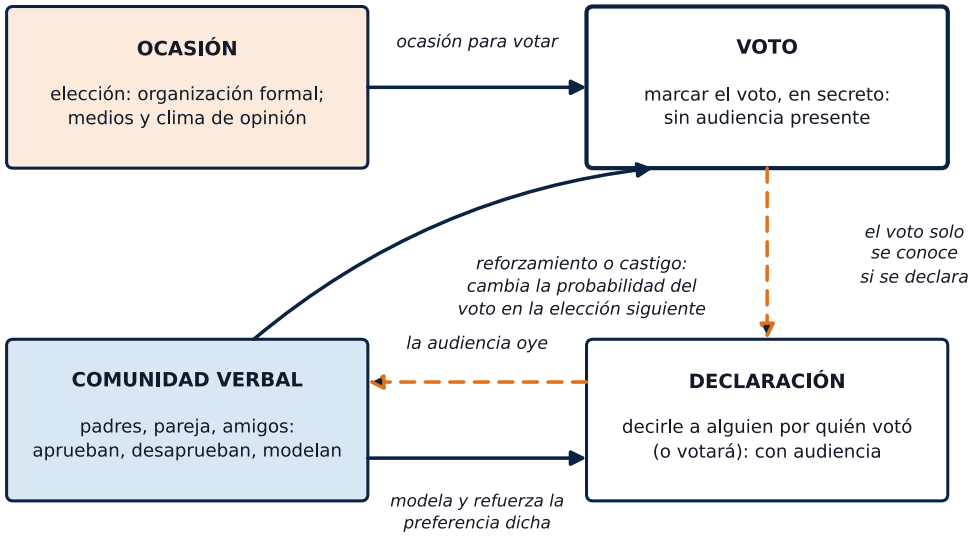
Visser (1996), cuyo artículo sirve de base a la sección, llega a la misma afirmación modesta por otro camino. El autor afirma que el voto rara vez se estudió desde la perspectiva analítico-conductual, en razón de la fuerte influencia mentalista en la investigación electoral. Parte de que votar por un partido es una respuesta operante, emitida bajo condiciones de estímulo especificadas: la organización formal de la elección es la ocasión; las consecuencias las aplica la comunidad verbal del elector (familia, amigos, colegas, vecinos); y fuentes distales, como el clima de la opinión pública, que los medios vuelven relevante para cada elector, también especifican probabilidades de reforzamiento o de desaprobación.

Para poner a prueba el modelo, Visser lo simplifica. Deja de lado los estímulos discriminativos de la situación electoral, suponiendo que basta que el elector sepa cuándo y dónde ocurre la elección y qué partidos compiten, y se queda con dos agentes de consecuencias: un ambiente social cercano (como la familia), que modela la conducta y refuerza la verbalización de una preferencia igual a la suya, y, en la vida adulta, los pares (una comunidad verbal más am-

plia: amigos, compañeros de trabajo, parejas), que refuerzan el voto semejante al suyo y desaprueban el divergente. La Figura 7.1 reúne el modelo y añade una caja que él no trae, la declaración.

FIGURA 7.1

El voto como episodio verbal mediado: ocasión, voto sin audiencia, declaración con audiencia y comunidad verbal



Continua: relación del modelo de Visser (1996). Discontinua: lectura aquí propuesta (la declaración no está en el modelo).

Nota. Esquema original del autor, a partir de Visser (1996) y de Skinner (1957). No representa datos. Línea continua: relación del modelo de Visser; discontinua: lectura aquí propuesta (la declaración no consta en el modelo).

La caja de la declaración es el añadido que hace posible el análisis, y Visser no lo hace. En el modelo, las consecuencias sociales inciden sobre el voto. Pero el voto es secreto, y lo que la comunidad del elector puede reforzar o castigar es lo que ve y oye: lo que el elector dice sobre su voto, antes o después, y lo que puede inferirse de lo que dice sobre otros asuntos. Entre el acto y la consecuencia social hay, en la mayoría de los casos, un enunciado. Es por ese enunciado que el voto entra en el episodio verbal, y es sobre él que trabaja el resto del capítulo.

7.2 ¿Mando, tacto o intraverbal? Funciones posibles del voto

Como en el Capítulo 3, lo que define al operante es lo que lo controla, y no la forma. Skinner (1957) afirma que el mando es un tipo de operante definido por sus variables de control, y no una unidad formal: ninguna respuesta puede

llamarse mando solo por la forma, e identificar cualquier operante verbal exige conocer las variables de las que la respuesta es función. Lo mismo vale para el tacto y para el intraverbal. Un mismo voto, por el mismo candidato, puede estar bajo controles diferentes en electores diferentes, y en el mismo elector en elecciones diferentes.

La primera candidata es la función de tacto. El tacto es el operante evocado por un objeto, evento o propiedad, porque en su presencia una respuesta de esa forma es reforzada de manera característica en la comunidad verbal (Skinner, 1957). Un voto bajo esa función estaría controlado por lo que el candidato o el partido hizo, dijo o representa para el elector, y variaría con el evento: si lo que lo controla cambiara, el voto cambiaría. Es la lectura más directa del voto. Su punto débil es que exige contacto con el evento. Guerin (1994) observa que muchas actitudes versan sobre eventos que la persona nunca experimentó, como la energía nuclear, y por eso serían intraverbales: lo que se sabe viene de los periódicos y de otras personas. Es razonable suponer que, para muchos electores, los candidatos a cargos nacionales sean de ese tipo, pero no hay datos sobre esto entre las fuentes leídas.

La segunda es la función de intraverbal, la respuesta sin correspondencia punto por punto con el estímulo verbal que la evoca, de la que, según Skinner (1957, pp. 71–72), está hecha la mayor parte de los "hechos" de la historia. Un voto bajo esa función estaría controlado por palabras: la sigla, el eslogan, lo que dice la familia, lo que el grupo recomienda. El modelo de Visser es, en este punto, casi una descripción de esa historia: la audiencia refuerza la verbalización de la preferencia semejante a la suya, de modo que, cuando llega la elección, el elector tiene una preferencia verbalizada que el acto solo continúa. En ese caso, el voto cambia cuando cambian las palabras que lo evocan y las personas que las dicen, y no cuando cambia el evento.

La tercera es la función de mando, o mejor, de algo que se le aproxima. El mando es la respuesta reforzada por una consecuencia característica y, por eso, bajo control de privación o de estimulación aversiva (Skinner, 1957). Votar para obtener o para evitar algo de quien gobierna es la lectura más natural, y la más frágil: el oyente no está presente, no fue entrenado para reforzar al elector, y el reforzador, si viene, viene a toda la comunidad y con gran retraso. La función de mando encaja mejor en la declaración que en el voto. Guerin (1994) muestra el mecanismo para los enunciados de actitud: la jefa que dice que no le gusta quien toma café en el trabajo no hace un tacto inocente de una actitud privada, y funciona como si hubiera mandado "¡No tome café mientras trabaja, o ya verá!", tenga o no esa intención. Quien dice "yo voto por fulano" ante alguien que aún no ha decidido puede estar haciendo eso, y la regla del Capítulo 5 vale aquí: la forma no decide.

Hay una cuarta posibilidad, que no es un operante verbal más, pero cambia lo que se puede esperar del voto: la regla. Street (1994), citando un texto de Craig de 1980, da el ejemplo del elector que dice ser demócrata y vota por un republicano para presidente. La investigación tradicional toma eso por inconsistencia de actitud; para Craig, puede ser solo un artefacto de la incapacidad del investigador para identificar aspectos salientes del estímulo, dada la historia de la persona, porque el elector puede tener una regla de nunca votar por un presidente del partido que controla el Congreso. La "inconsistencia" es, entonces, una medida de la incapacidad del observador para hacer sondeos verbales eficaces, dada la historia del hablante. El voto que se desprende de una regla es conducta gobernada por reglas, en los términos de la sección 3.7, y puede no ser ni tacto, ni intraverbal, ni mando.

Lo que se puede decir, en el estado de las fuentes, es poco. Las cuatro posibilidades son candidatas y ninguna se puso a prueba para el voto de un elector brasileño. El único modelo operante del voto entre las fuentes es el de Visser, que trata el voto como una respuesta mantenida por consecuencias sociales y deja de lado los estímulos discriminativos de la elección. Lo que el análisis gana con la lista es una pregunta precisa para cada caso: qué cambiaría el voto de este elector, y quién lo sabría.

7.3 El voto secreto como retirada de la audiencia

La audiencia es, para Skinner (1957), el estímulo discriminativo en cuya presencia la conducta verbal es reforzada de manera característica. Las audiencias difieren por lo que hacen con la conducta verbal: la que solo no refuerza se distingue de la que castiga, y una audiencia en cuya presencia la conducta verbal es castigada es una audiencia negativa. El confidente es el ejemplo de la audiencia no punitiva, la persona comprensiva a quien se le puede hablar con menos miedo al castigo que a oyentes cualesquiera, y, ante ella, el hablante puede emitir conducta que otras audiencias castigan. El Capítulo 10 (sección 10.1) trata de esas audiencias por entero; aquí bastan dos ideas: la audiencia es parte de la ocasión de la conducta verbal, y lo que una audiencia castiga puede emitirse ante otra.

El secreto de la cabina es, en estos términos, la retirada institucional de cualquier audiencia identificada en el momento del acto. El elector marca el voto sin que nadie que lo apruebe o lo repruebe esté presente, y sin que el acto, en ese momento, produzca una consecuencia social. La hipótesis del capítulo, que ninguna fuente verificada pone a prueba, es que eso desplaza el control: la audiencia deja de ser ocasión para el acto, y lo que queda es lo que la historia del elector dejó, incluidas reglas y preferencias verbalizadas, aprendidas ante

esas audiencias. Lo que la retirada no elimina es lo que el elector dijo antes y lo que dirá después. La declaración de voto, al contrario que el voto, se emite ante una audiencia, y es bajo su control donde se espera encontrarla.

Hay una consecuencia para quien investiga, que Visser (1996) anota con franqueza. Como el voto es secreto, el investigador no puede inspeccionarlo y recurre a relatos verbales del propio elector sobre su conducta de votar y sobre los factores que la refuerzan. Esos relatos pueden estar bajo control de contingencias diferentes de las del voto que supuestamente reflejan, en particular las de la situación de entrevista en que suelen recogerse los datos de las encuestas. Visser supone a continuación que, en una entrevista de encuesta que procura ser discreta, el relato refleja con precisión la conducta de votar y los factores que la controlan. Es una suposición de trabajo, declarada como tal, y el resto del artículo depende de ella.

Guerin (1994) proporciona el contrapunto. Las escalas y las preguntas de encuesta dependen de quién pregunta y de la cultura verbal en que se hace la encuesta, porque las actitudes dependen siempre de consecuencias sociales; y cita, en este punto, un estudio de 1986 sobre la presencia de otros y la declaración del voto en las elecciones estadounidenses. Ese estudio no se leyó; solo se conoce la referencia, y no se afirma lo que encontró. Para el capítulo, basta lo que Guerin extrae: el relato del voto, como otro relato, depende de quién pregunta.

DATO	Como el voto es secreto, el investigador no lo inspecciona y usa relatos verbales, que pueden estar bajo control de contingencias diferentes de las del voto, en particular las de la entrevista (Visser, 1996). Las escalas y las preguntas de encuesta dependen de quién pregunta y de la cultura verbal en que se hace la encuesta (Guerin, 1994).
INTERPRETACIÓN	Lo que la investigación electoral observa es la declaración del elector bajo la audiencia de la entrevista, y no el voto. La relación entre los dos es el objeto de estudio, y no puede presuponerse.
HIPÓTESIS	Que la declaración pública de voto esté más bajo control de la audiencia que el voto es una hipótesis razonable y no puesta a prueba: no hay, entre las fuentes verificadas, un estudio que compare, para los mismos electores, lo que declaran ante audiencias diferentes con el voto validado.

7.4 Declarar y votar: actitudes verbales y actos

La pregunta por la relación entre lo que la persona dice y lo que hace es más antigua que la investigación electoral, y el análisis de la conducta no es el único que la ha hecho. Lloyd (1994) parte del reconocimiento de sentido común de que hay una correlación baja, o nula, entre lo que la persona dice que haría y lo que de hecho hace después, y observa que el psicólogo, que se apoya con frecuencia en el relato verbal en lugar de observar la conducta criterio, es menos escéptico que el lego. La sección reúne lo que muestran las fuentes verificadas, y empieza por el concepto que la organiza.

CONCEPTO EN USO

Correspondencia decir-hacer

La correspondencia decir–hacer es la relación entre lo que la persona dice que hizo, o que hará, y lo que de hecho hace. En Risley y Hart (1968), doce niños de 4 a 5 años, de familias de bajos ingresos, en un preescolar experimental de medio tiempo, formaban dos grupos de seis. Las maestras registraban el uso de los materiales en la hora libre de la mañana (el hacer) y, cerca de una hora y media después, en la última merienda antes de que los niños se fueran a casa, le preguntaban a cada uno qué había hecho de bueno ese día (el decir). Primero, la merienda y la aprobación dependían de relatar el uso de un material específico: bloques en el grupo A, pintura en el grupo B. Con la merienda dependiendo solo del relato, la proporción de niños que relataban subió a una mediana del 100% en los dos grupos, sin efecto confiable sobre el uso. Después, la merienda pasó a depender de que el relato correspondiera al uso observado, y la correspondencia subió, por caminos diferentes: en el grupo A, el uso aumentó hasta el nivel del relato (mediana del 100% de decir y de hacer); en el grupo B, el relato cayó (a una mediana del 40%) hasta el nivel de quienes de hecho pintaron (25%).

Dos consecuencias importan aquí. La primera es que decir y hacer son, en palabras de Lloyd (1994), clases de respuestas separadas, cuyo desarrollo puede o no ser independiente, y no hay implicación de que una cause o influya en la otra. La segunda es que la correspondencia no es una propiedad de la persona: es el efecto de una contingencia que la refuerza. Reforzar el decir por sí solo no produce el hacer.

Risley y Hart partieron de estudios anteriores que reforzaron solo la conducta verbal y encontraron poco aumento de la conducta no verbal relacionada. El mismo estudio abre, en Lloyd (1994), la sección sobre hacer–decir y decir–hacer, y es por él que el concepto entra en la discusión sobre las actitudes. La Tabla 7.1 reúne cinco fuentes verificadas: dos, que el texto conoce por el relato

de Lloyd, muestran la distancia en un estudio de campo y en una revisión de 31 estudios; dos la miden ante un pedido concreto hecho al participante; y una muestra una condición que la reduce.

TABLA 7.1

Cinco fuentes sobre la distancia entre lo que se dice y lo que se hace, y una condición que la reduce

Fuente	"Decir"	"Hacer"	Resultado	Lo que no permite
LaPiere, en 1934, según el relato de Lloyd (1994)	Respuesta escrita de propietarios sobre atender a clientes chinos	Atención a una pareja china, en visitas a cerca de 250 establecimientos	Cerca del 90% de los 128 que respondieron dijeron que no los atenderían	Verificar el estudio original, que no se leyó
Wicker, en 1969, revisión relatada por Lloyd (1994)	Escalas de actitud	Medidas de conducta abierta, en 31 estudios	Correlaciones rara vez por encima de +0,30, muchas cerca de cero, negativas en cuatro	Verificar los 31 estudios; hablar del voto
DeFleur y Westie (1958)	Escala de diferencias sumadas; 23 estudiantes más y 23 menos prejuiciosos, emparejados	Firmar, en un formulario de usos graduados, el permiso para ser fotografiado con una persona negra del sexo opuesto	Los prejuiciosos evitaron más, pero 14 de 46 hicieron lo opuesto de lo previsto; el 71,8% invocó un grupo de referencia	Generalizar: 46 estudiantes; los autores no pusieron a prueba el mecanismo
Linn (1965)	Actitud verbal (prejuicio)	Posar para una fotografía con una persona negra del sexo opuesto	Discrepancia en el 59% de los casos; relación en función del involucramiento social y de la experiencia previa	Estimar la tasa en otras poblaciones
Regan y Fazio (1977)	Actitud, formada por experiencia directa o por información	Conducta relacionada, en laboratorio y en campo (crisis de vivienda)	Más consistencia con experiencia directa	Hablar de declaraciones públicas; del voto

Nota. Números de los artículos y, en las dos primeras filas, del relato de Lloyd (1994). Los estudios de LaPiere y de Wicker se conocen aquí solo por ese relato y, en el caso de LaPiere, por una mención en DeFleur y Westie (1958); las páginas citadas son las de Lloyd. Las "medidas de conducta" en Wicker son las que usaron los 31 estudios, y no el voto.

Las dos primeras filas dicen cuán antigua es la pregunta. LaPiere viajó por Estados Unidos con una pareja china, que fue atendida en más de 250 restaurantes, hoteles y establecimientos semejantes, con un único rechazo (DeFleur y Westie, 1958); cuando preguntó después, por escrito, si aceptarían clientes chinos, cerca del 90% de los 128 propietarios que respondieron dijeron que no (Lloyd, 1994). Wicker revisó 31 estudios que midieron la actitud con escalas y la conducta abierta: las correlaciones rara vez pasaron de +0,30, muchas quedaron cerca de cero y en cuatro estudios fueron negativas (Lloyd, 1994). Una

correlación de +0,30 es débil: elevada al cuadrado, indica que menos de una décima parte de la variación de una medida acompaña a la de la otra. El número que importa para el capítulo es el de Wicker: la distancia entre el decir y el hacer no es una excepción. La seriedad del número, escribe Lloyd, es mayor cuando se nota la aceptación difundida, entre legos y psicólogos, de la idea de que, para cambiar la conducta, primero hay que cambiar las actitudes, idea que exige una correlación alta y positiva.

Las dos filas siguientes son estudios en los que el hacer fue una decisión ante un pedido concreto, y lo que interesa es lo que los autores dicen del origen de la discrepancia. DeFleur y Westie (1958) aplicaron escalas de prejuicio a 250 estudiantes y eligieron a 23 del cuartil más prejuicioso y a 23 del menos prejuicioso, emparejados en ocho características. Después de una sesión de laboratorio y de una entrevista, pidieron que cada uno firmara, en un formulario de autorización con una serie graduada de usos, el permiso para ser fotografiado con una persona negra del sexo opuesto. Los prejuiciosos evitaron más, pero 14 de los 46 mostraron una conducta opuesta a la actitud verbal, y la relación, aunque significativa, no fue de uno a uno. En el 71,8% de los casos los sujetos invocaron algún grupo de referencia al decidir firmar o rechazar, y todos los que lo hicieron mencionaron algún grupo de pares; los autores atribuyeron la falta de relación lineal a restricciones sociales y sugirieron que las escalas de actitud tal vez necesiten una categorización sistemática de ellas.

Linn (1965) rehízo el diseño con una medida de actitud idéntica al acto. Treinta y cuatro universitarias de Wisconsin, de 18 y 19 años, habían respondido a un cuestionario con siete ítems del tipo “estaría dispuesta a posar con un negro del sexo opuesto si la fotografía fuera a usarse” en tal contexto, desde el uso en el laboratorio hasta un folleto de una campaña nacional por la integración racial. Cuatro semanas después, sin saber que participaban en un estudio, fueron invitadas a ayudar a construir un test de personalidad y firmaron o no las mismas siete autorizaciones de uso. En el 59% de los casos, la actitud verbal y el número de autorizaciones firmadas difirieron en dos o más puntos de siete. Linn concluyó que la relación entre prejuicio y discriminación es función del nivel de involucramiento social con el objeto y de la experiencia previa con él. En lenguaje operante, las restricciones sociales y los grupos de referencia son audiencia (secciones 7.1 a 7.3).

La condición que reduce la distancia, por su parte, es la experiencia directa. Regan y Fazio (1977) hicieron un experimento de laboratorio y un estudio de campo. En el laboratorio, 28 estudiantes de primer año voluntarios de Cornell fueron sorteados en dos condiciones: en la de experiencia directa, resolvieron los ejemplos de cinco tipos de problema; en la de experiencia indirecta, vieron los mismos ejemplos ya resueltos y explicados por el experimentador. To-

dos evaluaron el interés de cada tipo, de "extremadamente aburrido" a "extremadamente interesante", lo que midió la actitud, y después tuvieron 15 minutos libres para trabajar en los problemas que quisieran; el orden y la proporción intentada de cada tipo midieron la conducta. La consistencia entre actitud y conducta fue mayor en los sujetos cuya actitud se formó por experiencia conductual directa que en los de actitud formada por información del experimentador. En el estudio de campo, se enviaron cuestionarios a estudiantes de primer año a quienes la universidad había alojado en viviendas temporales por falta de plazas (experiencia directa de la crisis de vivienda) y a estudiantes de primer año alojados desde el inicio en viviendas permanentes (experiencia indirecta); midieron la actitud sobre la crisis en siete ítems y ofrecieron seis acciones posibles, como firmar una petición o escribir una carta, y la consistencia entre actitud y acción fue mayor en los estudiantes con experiencia directa. Para el voto, la pregunta es cuál sería la experiencia directa con el objeto: el candidato llega, para muchos electores, por palabras (sección 7.2), y la experiencia directa que tiene el elector es la de votar y la de declarar el voto.

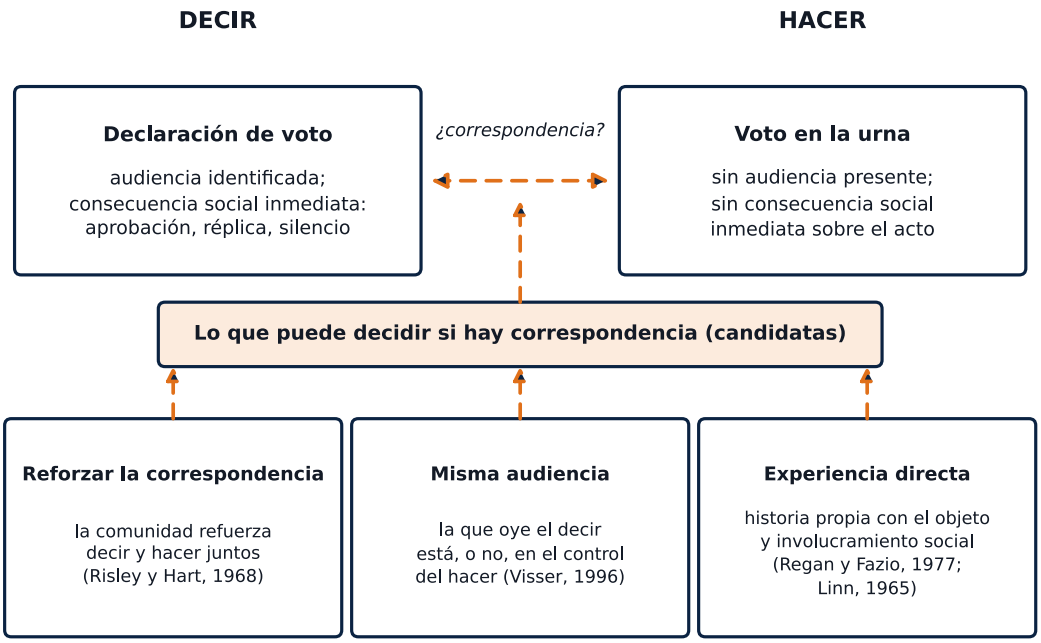
Para los analistas de la conducta, el problema no está solo en la medida. Street (1994) escribe que, en la visión analítico-conductual, las actitudes son las conductas que se ven cuando se le pide a alguien que llene una encuesta de actitud, sin que se esté captando un estado mental duradero, y que responder a una escala es una conducta emitida bajo condiciones sociales poco comunes. El cognitivista supone que el autorrelato y la conducta ante el objeto son efectos de la misma causa, la actitud; el analista, que cada uno es producido por sus propios eventos causales y que, si son diferentes, las conductas probablemente serán diferentes. El criterio de consistencia es, él mismo, verbal, y los autorrelatos son reforzados socialmente por su plausibilidad, y no por su exactitud objetiva. Guerin (1994) llega al mismo lugar: dadas las múltiples fuentes de control, los repertorios amplios y las varias comunidades verbales, deben esperarse inconsistencias entre actitud y conducta, sobre todo cuando la conducta es verbal.

La Figura 7.2 reduce la sección a una pregunta: ¿qué decide si lo que el elector dice y lo que hace se corresponden? Las tres candidatas salen de las fuentes: la historia de reforzamiento de la correspondencia (Risley y Hart, 1968); la identidad, o no, entre la audiencia del decir y la que controla el hacer, lectura que se desprende de Visser (1996) y de DeFleur y Westie (1958); y la experiencia directa con el objeto (Regan y Fazio, 1977; Linn, 1965). Ninguna se puso a prueba para la declaración de voto y el voto en el mismo elector.

Lloyd (1994) añade dos observaciones sobre el orden, ambas de estudios que no se leyeron. En un estudio de Israel y O'Leary, de 1973, los niños que jugaban durante 20 minutos y, 40 minutos después, decían a qué habían jugado (hacer–decir) mostraron mayor correspondencia que los que decían a qué jugarían y, 40 minutos después, jugaban (decir–hacer). Y, en un estudio de Sowers y colaboradores, de 1977, el relato de universitarios sobre el estudio hecho 14 horas antes correspondió más al estudio que lo que su planificación correspondió al estudio hecho 6 horas después. La lectura del capítulo es que declarar después del voto, como en una declaración recogida a la salida de la urna, es una respuesta del tipo hacer–decir, y declarar antes es del tipo decir–hacer.

FIGURA 7.2

Decir y hacer: lo que puede decidir si la declaración de voto y el voto se corresponden



Esquema original, sin datos. Discontinua: relación conceptual o hipótesis.

Nota. Esquema original del autor, a partir de Risley y Hart (1968), Visser (1996), Regan y Fazio (1977) y Linn (1965). No representa datos. Línea discontinua: relación conceptual o hipótesis.

DATO	En 31 estudios revisados por Wicker en 1969, las correlaciones entre escalas de actitud y medidas de conducta rara vez pasaron de +0,30 (Lloyd, 1994). En DeFleur y Westie (1958), 14 de 46 estudiantes actuaron en oposición a su actitud verbal; en Linn (1965), la discrepancia fue del 59%. En Risley y Hart (1968), reforzar el relato elevó el relato sin alterar el uso, y reforzar la correspondencia la produjo en los dos grupos, por caminos diferentes.
INTERPRETACIÓN	La distancia entre el decir y el hacer no es una anomalía que haya que explicar por error de medida: es lo que se espera de dos clases de respuestas con contingencias propias (Street, 1994; Guerin, 1994). Cuando hay correspondencia, parece depender de una historia de reforzamiento de la propia correspondencia (Risley y Hart, 1968) y de la experiencia directa con el objeto (Regan y Fazio, 1977). Nada en las fuentes trata del voto.
HIPÓTESIS	Que la correspondencia entre la declaración de voto y el voto sea mayor cuando la misma audiencia participa en los dos momentos, y menor cuando la audiencia de la declaración difiere de la que controla el voto, es una hipótesis aquí propuesta, y ninguna fuente la pone a prueba. Que declarar después de votar corresponda más que declarar antes, como relata Lloyd para niños y universitarios, puede ponerse a prueba con votos validados y declaraciones recogidas en momentos diferentes, y no hay, entre las fuentes leídas, un estudio que lo haga.

7.5 Consecuencias sociales de votar: los experimentos de campo

Si el análisis del voto como conducta verbal es correcto, lo que la comunidad del elector hace con el voto, o con lo que se dice de él, cambia la probabilidad de votar y de votar del mismo modo. Dos fuentes tratan de esto, con diseños opuestos: una encuesta, que solo puede asociar, y un experimento de campo, que manipuló.

La encuesta es la de Visser (1996), con datos del estudio electoral nacional holandés de 1986, una encuesta en dos olas, antes y justo después de la elección, y la única de la serie que midió el voto de las personas cercanas al elector. Visser clasifica a cada elector según la historia de reforzamiento que el modelo supone, es decir, según si los padres votan por el mismo partido y si hay una pareja cuyo partido se conoce, y verifica si el voto del elector coincide con el de la pareja o, sin pareja identificable, con el de los padres; en ese caso, la

dirección del voto se considera explicada por el modelo. Entre 1.255 electores, el modelo explica la dirección de 812 votos (65%), frente al 41% esperado si los electores se distribuyeran al azar entre las categorías. En cuanto a la consistencia, que Visser define como el número de veces que el elector elige el mismo partido en elecciones consecutivas, los votos modelados o reforzados por los padres o la pareja fueron más consistentes que los desaprobados por personas significativas, en general con una diferencia grande y significativa: el 68% (padres y pareja con el mismo voto) y el 67% (solo los padres) de los electores dijeron haber votado siempre por el mismo partido, frente a entre el 29% y el 36% en las celdas en que el elector divergía de los padres o de la pareja. La excepción que el autor destaca es la celda en que los padres y la pareja votaban diferente y el elector votaba como la pareja: 32%, lo que, para él, subraya la importancia de la historia de reforzamiento. Visser advierte que la investigación por encuesta carece del carácter controlado de las experimentales, de modo que las afirmaciones causales exigen cautela, y que el análisis al nivel del elector individual puede pasar por alto el impacto de desarrollos macrosociales; y concluye que, desde el punto de vista analítico-conductual, no hay razones válidas para conceder al voto una posición privilegiada entre las acciones humanas.

El experimento es el de Bond et al. (2012), con 61 millones de usuarios de Facebook en las elecciones legislativas estadounidenses de 2010. Los usuarios fueron asignados al azar a un mensaje social (60.055.176 usuarios: incentivo para votar, enlace al lugar de votación, botón “I Voted”, contador y hasta seis fotos de amigos que habían hecho clic en el botón), a un mensaje informativo (611.044: lo mismo, sin las fotos) o a ningún mensaje (613.096). Midieron tres cosas. Hacer clic en “I Voted” es una medida de autoexpresión política, que los autores dicen que probablemente se ve afectada por cuánto desea el usuario ser visto como votante por los demás; hacer clic en el enlace al lugar de votación mide la búsqueda de información, y no se informa a la comunidad social; y el voto se validó en registros públicos para 6,3 millones de usuarios.

Los resultados son tres. Los que recibieron el mensaje social hicieron más clic en “I Voted” que los del informativo (20,04% frente a 17,96%), y votaron un 0,39% más que los que no recibieron mensaje y un 0,39% más que los del informativo; la participación de los que recibieron el informativo fue idéntica a la de los controles. Por cada amigo cercano que recibió el mensaje social, el usuario fue un 0,224% más propenso a votar (voto validado) y un 0,099% a expresar su voto; no hubo evidencia de efecto de los amigos comunes sobre el voto validado ni sobre la búsqueda del lugar de votación. Y los autores estiman que el mensaje social aumentó la participación directamente en cerca de 60 mil votantes y, por contagio social, en otros 280 mil, en un total de 340 mil

votos, cerca del 0,14% de la población en edad de votar; todos los efectos son de intención de tratar, es decir, calculados sobre todos los asignados al mensaje, incluidos quienes ya habían votado anticipadamente o entraron demasiado tarde en Facebook para votar; por eso subestiman el efecto sobre quienes de hecho pudieron ser tratados. Casi toda la transmisión ocurrió entre amigos cercanos, que probablemente tienen una relación cara a cara.

En términos operantes, el mensaje social difiere del informativo en una sola cosa: las fotos de amigos que declararon haber votado. El experimento aísla, por lo tanto, que el dato social, y no la información sobre dónde y cuándo votar, añadió participación. Qué fue lo que actuó en el dato social, el diseño no lo separa. Tres lecturas son compatibles: la de modelo (ver a los amigos votando); la de audiencia (la anticipación de que los amigos verán el propio "I Voted", la autoexpresión que miden los autores); y la de regla (ver que personas del grupo votan vuelve aplicable una regla que el elector ya tenía). El decir y el hacer se midieron bajo la misma manipulación y variaron juntos, en el sentido de que el mensaje social aumentó los clics y el voto validado. Lo que el artículo no muestra es la correspondencia por usuario: el voto validado se estudió mediante un proceso a nivel de grupo, descrito en el material suplementario, y no se puede ver, para cada persona, si el clic correspondió al voto.

Hay otros experimentos de campo sobre presión social y participación electoral, y no se citan porque su texto completo no pudo leerse por vía legal. La laguna es del acervo de fuentes leídas, y no de la literatura. También por eso el capítulo no afirma que el efecto de Bond et al. se extienda a elecciones de otros países, a otras plataformas o a otros años.

DATO	En 1.255 electores holandeses, los votos modelados o reforzados por los padres o la pareja fueron más consistentes a lo largo de las elecciones que los desaprobados, y la dirección del voto se explicó en el 65% de los casos, frente al 41% esperado por azar (Visser, 1996). En 61 millones de usuarios estadounidenses, el mensaje con fotos de amigos aumentó el clic en "I Voted" y el voto validado (un 0,39% más que sin mensaje); el mensaje sin fotos no alteró la participación (Bond et al., 2012).
INTERPRETACIÓN	Los dos son compatibles con la lectura de que lo que la comunidad del elector hace con el voto, o con lo que se dice de él, cambia la probabilidad de votar y de votar del mismo modo. El de Bond et al. es de manipulación y de una plataforma; el de Visser es de asociación, y el propio autor advierte sobre la causalidad.

HIPÓTESIS Que el mecanismo sea la aprobación o la desaprobación anticipada por los amigos (reforzamiento social) es una hipótesis, y el diseño no la separa del modelo ni de la regla. Que lo mismo valga para las elecciones brasileñas es una extensión, y no un dato.

7.6 Nombres y números de candidatos en clases de equivalencia

Los nombres, siglas y números de los candidatos son los estímulos con los que el elector llega a la urna, y el Capítulo 4 (sección 4.5) dejó una hipótesis sobre ellos: que el rótulo derecha–izquierda forme parte de la clase con la que el nombre del político llega al elector. Lo que el Capítulo 4 tenía para sostenerla es poco. En el estudio de Pereira y Moreira (2021b), cinco estudiantes de Psicología aprendieron, por ensayo y error, qué figuras abstractas iban con cada una de tres fotografías de presidentes de Brasil (Bolsonaro, Lula y Fernando Henrique Cardoso). Las siglas de los partidos (PSL, Partido Social Liberal; PT, Partido dos Trabalhadores; y PSDB, Partido da Social Democracia Brasileira) y las palabras derecha, izquierda y centro se relacionaron correctamente con las figuras abstractas, sin entrenamiento específico, lo que los autores toman como indicio de que ya formaban parte, antes del experimento, de una clase con las fotografías de los presidentes; y, después de aprender las relaciones entre fotografías y figuras abstractas, las evaluaciones de las figuras, hechas en una escala de adjetivos opuestos, quedaron más próximas a las de la fotografía correspondiente. En otro estudio de los mismos autores, con dos experimentos de cinco participantes cada uno, el aumento de la semejanza entre las evaluaciones (100 menos la diferencia absoluta entre ellas) fue, con caras neutras de personas desconocidas, la mitad del obtenido con figuras abstractas, lo que atribuyen, probablemente, a la carga afectiva previa de las caras (Pereira y Moreira, 2021a). Los autores dicen no conocer estudios de actitudes con estímulos de significado social vinculados a posiciones políticas.

Hay una salvedad de resultado, que el Capítulo 4 ya dio. Haydu et al. (2015) enseñaron, por igualación a la muestra, a 28 hombres de barras organizadas de tres clubes brasileños a relacionar los escudos de los clubes con pinturas abstractas y esas pinturas con las palabras bueno, malo y regular. Las pruebas evaluaban si emergían, sin enseñanza directa, relaciones entre el escudo del club apoyado y la palabra malo, entre el del principal rival y la palabra bueno y entre el del rival menor y la palabra regular, de modo que las palabras entraban en conflicto con la historia de cada hinchada. Los participantes formaron las relaciones enseñadas (de línea base) y las de simetría, pero no las de transitividad y de equivalencia entre escudos y palabras, y los errores más frecuentes fueron elegir lo que la historia prescribía: bueno para el escudo del

club apoyado y malo para el del principal rival. Lo que el estudio muestra, para este capítulo, es que la clase pretendida no siempre se forma cuando contraría la historia del participante. No se generaliza el hallazgo a la afiliación partidaria, y tampoco el de Pereira y Moreira.

La hipótesis del capítulo es la siguiente. En cualquier elección en que escoger exija reconocer o producir un nombre, una sigla, una imagen o un número de candidato, el acto supone una discriminación condicional entre esos estímulos, y el análisis de clases de equivalencia trata de estímulos arbitrariamente relacionados por la historia, que se vuelven sustituibles, con la función de uno llegando a los otros sin contacto directo (sección 4.3). Si el nombre, la imagen, la sigla y el número del candidato forman parte de una misma clase para el elector, la función evaluativa de cualquiera de ellos puede llegar a los demás, y el voto puede estar menos bajo control de lo que el candidato dijo (Capítulos 3 y 5) que de las relaciones que el elector aprendió entre los estímulos que lo designan. No hay fuente leída sobre el procedimiento de votación adoptado en Brasil, y por eso no se describe: la hipótesis vale para cualquier procedimiento que le pida al elector uno de esos estímulos.

Lo que pondría a prueba la hipótesis es un estudio que midiera, para candidatos de campos políticos diferentes y con el mismo procedimiento para todos, las relaciones entre nombre, imagen, sigla y número (por igualación a la muestra o por una tarea de denominación), la función evaluativa de cada uno (por diferencial semántico, como en el Capítulo 4) y la transferencia de esa función después de una historia de emparejamiento. La regla de simetría vale: mismo procedimiento, mismos criterios, para todos los campos.

DATO	Las siglas de partidos y las palabras derecha, izquierda y centro se relacionaron correctamente con figuras abstractas, sin entrenamiento específico, en cinco estudiantes (Pereira y Moreira, 2021b). En 28 hinchas, las relaciones de línea base y de simetría se formaron, pero no la transitividad ni la equivalencia entre escudos y palabras (Haydu et al., 2015).
INTERPRETACIÓN	La transferencia de función evaluativa entre fotografías, siglas y figuras es posible en el laboratorio y depende de la historia previa con los estímulos; donde lo que se enseña contraría la historia del participante, la clase pretendida no emergió (Haydu et al., 2015).

HIPÓTESIS Que el número del candidato, como el nombre y la sigla, entre en una clase con la fotografía y con el rótulo derecha–izquierda, y que eso medie el voto, es una hipótesis aquí propuesta. Ninguna fuente verificada manipuló números de candidatos ni midió el voto en estudios de equivalencia con electores.

7.7 Interpretaciones rivales y límites

La lectura del capítulo, de que el voto es una conducta cuyo mantenimiento es social y cuya relación con lo que el elector dice depende de contingencias propias de cada uno, tiene rivales, y la Tabla 7.2 las reúne con lo que las separaría de ella.

Los límites del capítulo son los de las fuentes. Ninguna mide el voto de electores brasileños: el estudio del voto en términos operantes es una encuesta holandesa de 1986, sobre cuya causalidad el propio autor hace salvedades, y el experimento es de una plataforma, en un país, en una elección. La literatura de actitud y conducta mide la actitud, y no el voto, y LaPiere y Wicker se conocen aquí solo por el relato de Lloyd. La clasificación del voto como conducta verbal se propone aquí, y vale como análisis, y no como hallazgo. Y ninguna fuente compara campos políticos, de modo que el capítulo no dice nada sobre diferencias entre electores de campos diferentes.

TABLA 7.2

Lecturas rivales del voto y de lo que el elector dice de él: lo que las sostendría y lo que las separaría

Lectura rival	Lo que la sostendría	Lo que la separaría de la lectura del capítulo
El voto es conducta no verbal, sin mediación social; la declaración es otra respuesta y la relación entre ellas no es una pregunta sobre el voto	Un voto que no varía cuando cambian las consecuencias sociales, ni después de declaraciones	El mensaje social de Bond et al. cambió el voto validado; el modelo de Visser explica el 65% de los votos
Ambiente compartido: los padres y la pareja votan igual porque comparten información, barrio e ingresos, y la consistencia se desprende de eso, y no del reforzamiento (Visser, 1996, sobre la causalidad)	Consistencia igual entre electores que no son aprobados ni desaprobados por los padres o la pareja	Un diseño que mantenga el ambiente y varíe la aprobación, o que siga al elector en el tiempo
Actitud como disposición: la declaración y el voto son efectos de la misma causa, y las inconsistencias vienen de una mala medida (descrito por Street, 1994)	Correlaciones altas entre la declaración y el voto validado con mejores instrumentos	Decir y hacer variando por razones diferentes (Risley y Hart, 1968; DeFleur y Westie, 1958)
Recordatorio e información: el mensaje social solo recordó que había que votar	Efecto igual con el mensaje informativo	Bond et al.: el informativo tuvo la misma participación que el control; pero el experimento no separa modelo, audiencia y regla
Relato impreciso: la consistencia de Visser viene de la memoria y de la deseabilidad, y no de la conducta	Un relato de votos pasados que se aproxima al del voto presente	Voto validado en un registro público, como en Bond et al.

Síntesis

Votar es una conducta cuyo mantenimiento es social, y por eso le corresponde el análisis de la conducta verbal, aunque el oyente del acto sea un sistema, y no una persona presente (7.1). Un mismo voto puede estar bajo control de tacto, de intraverbal, de mando o de regla, y la forma no decide (7.2). El secreto le quita al acto la audiencia identificada y deja la declaración, que la tiene; lo que la investigación electoral observa es la declaración, bajo la audiencia de la entrevista, y no el voto (7.3). La relación entre decir y hacer es, desde LaPiere, baja y variable, y el análisis de la conducta la lee como la de dos clases de respuestas con contingencias propias; la correspondencia es el efecto de una contingencia que la refuerza, y reforzar el decir por sí solo no produce el hacer (7.4). Hay datos de asociación y datos de manipulación de que lo que la comu-

nidad hace cambia la probabilidad de votar, pero el diseño no separa modelo, audiencia y regla (7.5). La hipótesis de que el nombre, la imagen, la sigla y el número del candidato formen una clase para el elector se propone aquí, y ninguna fuente la pone a prueba (7.6).

El Capítulo 10 (sección 10.1) trata por entero de las audiencias negativa y no punitiva de las que se sirve la sección 7.3, y el 12 retoma la correspondencia decir–hacer al evaluar intervenciones cuyo desenlace es lo que el participante dice.

Referencias

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* (1.ª 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

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

instituto WALDEN4
walden4.com.br



PARTE III

La ecología de la información

8 Economía de la atención: el algoritmo como programador de contingencias

9 El ciclo productor-audiencia

10 La ventana de Overton: lo que se puede decir sin ser castigado

11 Infodemiología: cómo se propaga la información y qué se mide

CAPÍTULO

8

Economía de la atención: el algoritmo como programador de contingencias

Upworthy, un sitio de noticias virales, ponía a prueba titulares en experimentos, y uno de ellos sirve de punto de partida, aunque será mejor abordado más adelante. Una misma noticia se presentó a los lectores con cuatro titulares diferentes, y las tasas de clics fueron del 0,94%, el 0,79%, el 0,60% y el 0,58% (Robertson et al., 2023, Tabla 2). Con la misma nota y el mismo medio, el titular más cliqueado recibió cerca de 1,6 veces los clics del menos cliqueado (cálculo propio). El sitio lo hacía en serie: el conjunto de datos reúne 22.743 pruebas aleatorizadas, con cerca de 105 mil variaciones de titular.

Cada prueba es un arreglo experimental limpio: el titular varía, la noticia y el público son los mismos, y lo que se mide es la atención que obtiene cada variación, contada en clics. ¿Qué ocurre con lo que se dice cuando es así como se distribuye la atención? El capítulo trata de la economía en la que se produce esa distribución. La sección 8.1 presenta la atención como reforzador; la 8.2, las propiedades del contenido que los datos asocian con más clics, más veces compartido y más reacciones; la 8.3, la regla con la que la plataforma decide cuántos ven cada emisión; la 8.4, la hipótesis de que esa entrega funcione como reforzamiento intermitente; y la 8.5, los experimentos en los que se manipuló la regla.

El capítulo es cauteloso en dos puntos que la conversación corriente da por resueltos. El primero es que el feed sea un programa de razón variable: ninguna fuente verificada mide programa alguno, y la sección 8.4 dice lo que describe la literatura sobre los programas y lo que faltaría para la comparación. El segundo es que el algoritmo premie la ira o favorezca a un campo político: lo

que muestran los datos es más estrecho, y la sección 8.3 lo describe en los términos de los autores. En todas las secciones, dato, interpretación e hipótesis quedan separados.

8.1 La atención como reforzador

Skinner (1957) trata la conducta verbal como conducta mantenida por consecuencias mediadas por oyentes, y la atención del oyente aparece en pasajes dispersos del libro, el más explícito en la p. 199. El recuadro reúne lo que él dice y lo que este capítulo toma de allí.

CONCEPTO EN USO

La atención como reforzador condicionado generalizado (Skinner, 1957)

El Capítulo 3 (sección 3.3) presentó el reforzador condicionado generalizado: el evento que suele preceder a muchos reforzadores diferentes y que, por eso, puede reforzar la conducta bajo todas las condiciones apropiadas de privación y de estimulación aversiva, como la aprobación (Skinner, 1957). La atención del oyente tiene la misma estructura: es la consecuencia común de mandos generalizados como *¿Cómo?* y *¡Dónde!*, y, como la conducta siguiente del oyente puede ser relevante para muchos estados de privación, esas respuestas quedan bajo un control amplio. Skinner la da como ejemplo de variable que fortalece la conducta verbal con independencia de la forma: cualquier conducta verbal que evoque atención es reforzada, aparte de otras acciones específicas del oyente. El reforzamiento social generalizado de Guerin (1994), que el Capítulo 3 también presentó, es un concepto vecino: las actitudes son enunciados mantenidos por reforzamiento social generalizado de una comunidad verbal.

Dos puntos del recuadro importan para el resto del capítulo. El primero es la razón por la que la atención es un reforzador de amplio alcance: es condición para que el oyente haga cualquier otra cosa por quien habla, de modo que sirve a muchos estados de privación al mismo tiempo. El segundo es la cláusula "con independencia de la forma". Lo que se fortalece es cualquier respuesta que evoque atención, y la atención, como consecuencia, no distingue por sí misma lo verdadero de lo falso ni lo que informa de lo que irrita. Esta última frase es una lectura aquí propuesta a partir de Skinner, que no trata de desinformación ni de plataformas. Es ella, sin embargo, la que da sentido a preguntar, en las secciones siguientes, qué propiedades del contenido atraen atención y quién controla su distribución.

En una plataforma, la atención del oyente no es observable por quien habla. El productor ve recuentos: número de seguidores, "me gusta", veces compartido, comentarios, clics. La medida más fina entre las fuentes verificadas es la "impresión de permanencia" de Twitter, el registro de que al menos el 50% del

área de una publicación quedó visible durante al menos 500 milisegundos, que Huszár et al. (2022) llaman la mejor aproximación disponible para saber si el usuario fue expuesto al contenido. Es una medida de exposición, y no de atención en sentido funcional; el resto de lo que traen las fuentes son reacciones explícitas. Este capítulo llama atención medida al recuento de reacciones que la plataforma devuelve. La distancia entre ella y la atención que refuerza, la del oyente, es una premisa de lectura, y no un hallazgo.

Hay un estudio que pregunta directamente cuánto vale la atención entre las razones para usar redes sociales. Gunnarsson et al. (2026) aplicaron a 328 estudiantes de una universidad canadiense (edad media de 22,7 años; 68% mujeres) una tarea hipotética de compra de redes sociales. Los participantes imaginaban que solo podrían usar las redes comprando el tiempo, vendido por minuto, y decían cuántos minutos comprarían por día a cada uno de 15 precios, de 0,01 a 1.120 dólares por minuto, en cuatro condiciones de función: usar las redes únicamente para escapar del aburrimiento y de sentimientos desagradables (escape), para buscar información (tangible), para obtener interacción con el propio contenido, como "me gusta", veces compartido y comentarios (atención), o para sentirse bien (automática). La función afectó el consumo a costo mínimo (Q_0, lo que se compraría si el tiempo costara cero), pero no la sensibilidad al precio (α, la rapidez con la que la compra cae cuando el precio sube), y, en la formulación del resumen de los autores, el escape, la información o el reforzamiento positivo automático generaron más demanda que buscar "me gusta". En la condición de atención, el 57,6% de las trayectorias individuales no cumplió el criterio de sistematicidad ΔQ, uno de los indicadores de calidad de los datos de demanda (13,4%, 20,7% y 11,6% en las demás condiciones), en su mayoría por consumo cero, y el 51,5% de los participantes registró solo valores cero (7,0% en la condición tangible). Los autores leen esto como baja demanda de usar las redes sociales para obtener atención, y no como desatención a la tarea, pero advierten que la tarea es hipotética.

Lo que el estudio le dice al capítulo es una cautela. Si la atención fuera, para quien usa redes, el reforzador dominante, se esperaría una demanda alta justamente en la condición que la nombra, y no fue lo que apareció, en una muestra de estudiantes y en una tarea que aisló una función, aunque muchas personas usan las redes por funciones múltiples y superpuestas (limitación declarada por los autores). Hay además una diferencia de constructo: el estudio mide el valor, para el usuario, de usar la red para obtener atención, y no la eficacia de la atención como consecuencia de una publicación específica, que es la pregunta del capítulo. Las dos cosas pueden divergir.

DATO	Skinner (1957) describe la atención del oyente como consecuencia común de mandos generalizados y como variable que fortalece la conducta verbal con independencia de la forma; en una tarea hipotética con estudiantes universitarios canadienses, la demanda de usar las redes para obtener atención fue menor que la de escape, información o reforzamiento positivo automático (Gunnarsson et al., 2026).
INTERPRETACIÓN	La atención del oyente es un reforzador generalizado de amplio alcance para la conducta verbal. En una plataforma, lo que llega al productor es un recuento de reacciones, que sirve de medida aproximada de esa atención.
HIPÓTESIS	El recuento de reacciones es el reforzador que mantiene lo que el productor publica y es dominante en relación con otros. Los estudios del Capítulo 9 dan apoyo parcial a la primera mitad ("me gusta" y frecuencia de publicar; retorno social e indignación), con efectos pequeños y poblaciones específicas; la segunda mitad no se ha mostrado, y el resultado de Gunnarsson et al. (2026) pide cautela.

8.2 Lo que muestran los datos sobre el contenido emocional, moralizado y negativo

Tres estudios verificados vinculan propiedades del contenido con la reacción de la audiencia. Miden respuestas diferentes (hacer clic, compartir, publicar), y la diferencia es parte de lo que hay que decir.

Robertson et al. (2023) analizaron las pruebas de Upworthy, en las que cada visitante del sitio veía al azar una de varias variaciones de titular de una misma noticia, y se registraba el clic para leer la nota (visitantes entre 2013 y 2015). El conjunto analizado reúne 22.743 pruebas, de las cuales 12.448 pasaron el filtro prerregistrado, que descarta, por ejemplo, las pruebas con un único titular (muchas ponían a prueba imágenes de presentación, y no titulares). Esas pruebas compararon en promedio 4,31 variaciones cada una, para un total de 53.699 titulares, más de 205 millones de impresiones y 2.778.124 clics; la tasa media fue del 1,39%. Las palabras de los titulares se clasificaron, una por una, con un diccionario de lenguaje, el LIWC, con 620 palabras positivas y 744 negativas. Para un titular de longitud media, cada palabra negativa adicional aumentó la tasa de clics en un 2,3%, y la presencia de palabras positivas la redujo en cerca de un 1,0%. Los autores destacan que las palabras positivas eran algo más frecuentes que las negativas, el 2,83% frente al 2,62% de las palabras. El efecto de las palabras negativas fue más fuerte en las noticias de Gobierno y

Economía. Entre las emociones discretas, la tristeza aumentó la probabilidad de clic y la alegría la redujo; la ira no tuvo un efecto estadísticamente significativo, y el miedo tuvo un efecto en sentido opuesto al prerregistrado.

El diseño tiene límites que los propios autores declaran. Upworthy usaba titulares de clickbait y difiere de las fuentes tradicionales de noticias; las conclusiones son del nivel de las noticias, y no de los individuos; y nada dicen de la experiencia subjetiva de los lectores. Los titulares recibían menos clics cuando se publicaban más tarde en la historia del sitio, lo que los autores leen como indicio de que el clickbait era más eficaz cuando era nuevo para los lectores. Además, lo que el estudio mide es consumo, que ellos describen como conducta privada, a diferencia del compartir, que es público y está influido por factores sociales, como señalar identidad de grupo y mantener la reputación; eso puede explicar que el miedo y la ira pesen más en la decisión de compartir. La explicación es de los autores y no se puso a prueba con los datos del artículo.

Los otros dos estudios aparecen también en el Capítulo 9 y entran aquí solo para la comparación. Rathje et al. (2021) contaron, en 2.730.215 publicaciones de medios de noticias y de congresistas de los Estados Unidos en Facebook y en Twitter, las palabras que se referían al campo político opuesto (el exogrupo) y al propio campo (el endogrupo), y relacionaron esos recuentos con el número de veces compartido. Las publicaciones que se referían al exogrupo se compartieron cerca de dos veces más, y cada término sobre el exogrupo se asoció, en promedio, con un aumento de cerca del 67% (que los autores describen como aumento de la probabilidad de compartir), un efecto 4,8 veces el del lenguaje de afecto negativo y 6,7 veces el del lenguaje moral-emocional (sección 9.3). Brady et al. (2021) estudiaron la expresión de indignación del propio productor. En 12,7 millones de tuits de 7.331 usuarios, un aumento del 100% en el retorno social (“me gusta” y veces compartido) a la indignación del día anterior se asoció con un aumento esperado del 2 al 3% en la indignación del día siguiente; en dos experimentos de 120 participantes, que elegían tuits para publicar en una red simulada, la elección de tuits de indignación creció con el retorno y siguió la norma exhibida en el feed (sección 9.5).

La comparación enseña más por lo que impide que por lo que permite. “El contenido negativo es más reforzado” no es una proposición que describa un proceso único; son al menos tres: los titulares con palabras negativas reciben más clics (Robertson et al.), las publicaciones que mencionan al exogrupo se comparten más (Rathje et al.) y la expresión de indignación sigue el retorno recibido (Brady et al.). Son respuestas distintas, de personas en posiciones distintas del ciclo que se verá en el Capítulo 9, y nada obliga a que la misma consecuencia las mantenga. El patrón de las emociones discretas muestra la dife-

rencia: la ira, que la idea corriente pone como motor de la viralidad, no tuvo un efecto significativo en los clics, mientras que el lenguaje sobre el exogrupo predijo reacciones "angry" en las publicaciones de Facebook (Rathje et al.). La explicación de los autores de que consumir es privado y compartir es público es una hipótesis funcional, y el capítulo la toma con ese nombre.

TABLA 8.1

Tres respuestas al contenido emocional y moralizado: lo que midió cada estudio y lo que no permite decir

Estudio	Respuesta y material	Propiedad del contenido	Lo que observaron	Salvedad
Robertson et al. (2023)	Hacer clic en el titular (consumo, conducta privada). 12.448 pruebas aleatorizadas de Upworthy; 53.699 titulares; más de 205 millones de impresiones	Palabras negativas y positivas y emociones discretas, por diccionario	Cada palabra negativa: +2,3% de clics (titular de longitud media); positivas: cerca de –1,0%; la tristeza aumentó, la alegría disminuyó; ira sin efecto significativo	Medio de clickbait; nivel de la noticia, no del individuo; nada sobre la experiencia del lector
Rathje et al. (2021)	Compartir y reaccionar (conducta pública). 2.730.215 publicaciones de medios de noticias y de congresistas de los EE. UU., en Facebook y en Twitter	Referencia al exogrupo político; lenguaje de afecto negativo y moral-emocional	Exogrupo: cerca de dos veces más compartido; +67% por término, 4,8 y 6,7 veces el efecto del afecto negativo y del moral-emocional; reacción "angry" más popular en siete de ocho conjuntos	Correlacional; periodo anterior a la elección de 2020; no está claro cómo contribuyen las elecciones algorítmicas
Brady et al. (2021)	Publicar indignación (conducta del productor). 12,7 millones de tuits de 7.331 usuarios; dos experimentos de 120 participantes	Indignación moral expresada, medida por clasificador	+2 a 3% de indignación al día siguiente por cada +100% de retorno; en los experimentos, la elección de indignación creció con el retorno y siguió la norma	Los estudios observacionales no permiten inferencia causal; usuarios identificados por episodios de indignación pública; mide texto, no emoción

Nota. Los números vienen de los artículos. Las páginas de Robertson et al. (2023) son las de Nature Human Behaviour; Rathje et al. (2021) y Brady et al. (2021) no tienen paginación de revista. La columna "Propiedad del contenido" describe forma (palabras, referencia a un grupo, expresión clasificada); la función que la forma cumple para quien hace clic, comparte o publica no se midió en ninguno de los tres.

Está además la diferencia entre forma y función. Las propiedades medidas son de forma: palabras de un diccionario, la referencia a un grupo, una expresión de indignación clasificada por un modelo semántico. El análisis operante pregunta por la función, es decir, qué hace de una palabra negativa en el titular una ocasión para hacer clic, o de una mención al exogrupo una ocasión para reenviar. En los autores aparecen dos respuestas, señalar identidad de grupo y mantener la reputación (Robertson et al.) y satisfacer motivos de identidad social (Rathje et al.), y ambas son funciones propuestas que ningún estudio manipula. Sin la función, "contenido negativo" es topografía, y el tratamiento por función presentado en el Capítulo 3 muestra que muchas veces una misma forma presenta funciones distintas o está bajo control múltiple.

DATO	En pruebas aleatorizadas con más de 2,7 millones de clics, las palabras negativas aumentaron la tasa de clics y las palabras positivas la redujeron (Robertson et al., 2023); las publicaciones sobre el exogrupo político se comparten cerca de dos veces más que las publicaciones sobre el endogrupo (Rathje et al., 2021); la indignación del día siguiente sigue el retorno recibido por la del día anterior, con un efecto pequeño (Brady et al., 2021).
INTERPRETACIÓN	Propiedades de forma del contenido (palabras, referencia al exogrupo) se asocian con más clics, más veces compartido y más reacciones, y la asociación difiere según la respuesta medida, lo que sugiere operantes distintos con consecuencias distintas.
HIPÓTESIS	Lo que hace del contenido moralizado una ocasión para reenviar es la función que cumple para quien reenvía, por ejemplo señalar identidad de grupo o mantener la reputación (lectura de los autores). Ninguno de los estudios manipula la función.

8.3 El algoritmo como programador: cuánta audiencia recibe cada emisión

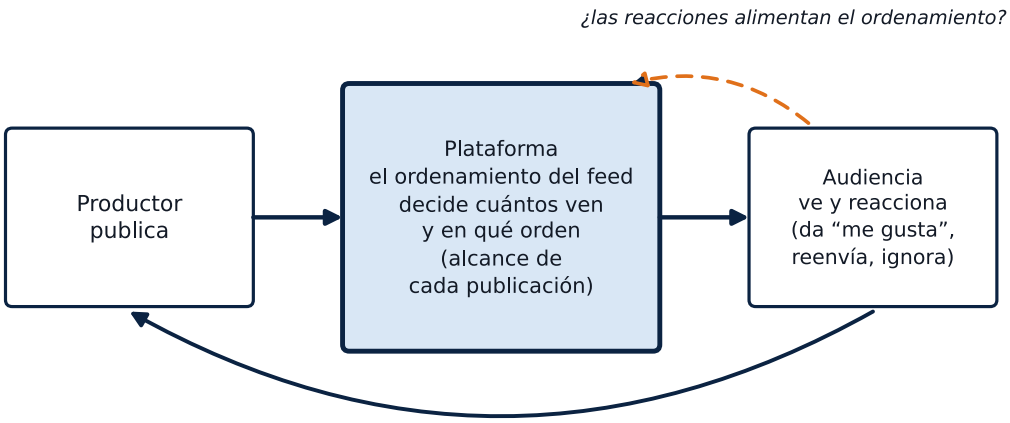
En las secciones anteriores, el productor publica y la audiencia reacciona; falta el término que decide cuántos ven cada emisión y en qué orden, y de él trata esta sección. En una plataforma con un feed ordenado, ese término es un conjunto de modelos de aprendizaje automático; en términos operantes, es la regla que fija, para cada emisión, el tamaño de la audiencia que podrá reaccionar a ella. Skinner (1957) ya notaba que los efectos debilitantes del reforzamiento intermitente o demorado se compensan en parte, en la conducta verbal, por el hecho de que los efectos de la conducta verbal pueden multiplicar-

se al exponer muchos oídos a las mismas ondas sonoras o muchos ojos a la misma página. La plataforma controla esa multiplicación: el ordenamiento decide cuántos ojos accederán a ella y reaccionarán.

La Figura 8.1 es un modelo, y la nota dice dónde hay datos. Las flechas continuas entre productor, plataforma y audiencia tienen apoyo en la auditoría de Huszár et al. (2022), descrita a continuación, y la que lleva los recuentos de vuelta al productor, en los estudios del Capítulo 9. La discontinua, en la que las reacciones alimentan el ordenamiento, es hipótesis. Brady et al. (2021) afirman que los algoritmos del feed, al determinar cuántos usuarios ven una publicación, afectan el retorno social y pueden influir en conductas morales explotando tendencias de aprendizaje por reforzamiento; es interpretación de los autores, y el estudio no mide el algoritmo.

FIGURA 8.1

El algoritmo como programador de las contingencias del feed



Línea continua: relación con datos (alcance con ordenamiento y con orden cronológico, Huszár et al., 2022; sensibilidad de quien publica al retorno, Lindström et al., 2021). Discontinua: hipótesis; lo que el ordenamiento usa como entrada no se conoce en las fuentes del capítulo.

Nota. Esquema original del autor, sin datos. Línea continua: relación con datos (alcance con ordenamiento y con orden cronológico, Huszár et al., 2022; sensibilidad de quien publica al retorno, Lindström et al., 2021). Discontinua: hipótesis; lo que el ordenamiento usa como entrada no se conoce en las fuentes del capítulo.

Huszár et al. (2022) son la fuente verificada que compara el ordenamiento personalizado con el orden cronológico del algoritmo sobre la audiencia. En 2016 Twitter introdujo algoritmos de aprendizaje automático para ordenar la Home timeline según un modelo de relevancia personalizada; antes, esta mostraba, en orden cronológico inverso, los tuits de las cuentas seguidas. Al introducir

la personalización, excluyó de ella a un grupo de control del 1% de los usuarios del mundo, que siguió con la línea de tiempo cronológica; el tratamiento era una muestra del 4% de las demás cuentas. El 5 de junio de 2020 el experimento incluía 58.087.969 identificadores de cuenta, el 5% de las cuentas del mundo, el 20% en el control y el 80% en el tratamiento. La amplificación es la razón entre el alcance de un conjunto de tuits en el grupo de tratamiento y en el de control: el 0% equivale a un alcance igual, y los valores por encima del 200% indican una audiencia tres veces la de la línea de tiempo cronológica. El alcance se contó en impresiones de permanencia (sección 8.1).

Todos los partidos analizados tuvieron una amplificación positiva por la personalización, en algunos casos por encima del 200%. Al comparar el mayor partido de derecha o centroderecha y el mayor de izquierda o centroizquierda de cada país, con excepción de Alemania, el primero tuvo una amplificación significativamente mayor; el efecto fue más fuerte en Canadá (liberales 43%, conservadores 167%) y en el Reino Unido (laboristas 112%, conservadores 176%). Los siete países son Estados Unidos, Japón, el Reino Unido, Francia, España, Canadá y Alemania; Brasil no está entre ellos.

El resultado cambia de naturaleza cuando la unidad es el político, y no el partido. En el nivel individual, la amplificación varió mucho dentro de cada partido, de cerca del 400% a menos del 0%, y una prueba de permutación (que compara el resultado con el obtenido al barajar al azar los partidos entre los políticos) no detectó una asociación estadísticamente significativa entre el partido y la amplificación de cada político. Los autores no ven contradicción: políticos diferentes pueden alcanzar audiencias superpuestas, de modo que pueden surgir correlaciones de grupo aun sin correlación individual. En los países con partidos de extrema izquierda o de extrema derecha bien representados, la amplificación de esos partidos fue en general menor que la de los moderados, y el análisis no sostiene la hipótesis de que la personalización amplifique las ideologías extremas más que las voces moderadas; los autores advierten que un fuerte sesgo partidario en las noticias se asoció con una mayor amplificación, lo que no implica promoción de ideología extrema. Tampoco hubo un patrón consistente entre gobierno y oposición: en el Reino Unido la amplificación favoreció a los conservadores en el gobierno y, en Canadá, a los conservadores de la oposición.

Para las noticias, el análisis usó 100.575.284 tuits que apuntaban a 6.258.032 artículos, clasificados por AllSides, una base que etiqueta la posición política de cada medio en una escala de cinco puntos, de la izquierda partidaria a la derecha partidaria: las fuentes más partidarias fueron más amplificadas que las de centro, y la derecha partidaria algo más que la izquierda partidaria. Con

la base Ad Fontes, otra base independiente de etiquetas, la izquierda partidaria tuvo una amplificación baja (10,5%), y los autores dicen que el resultado depende de la base de etiquetas políticas.

Las limitaciones vienen declaradas. El grupo de control no está aislado de efectos indirectos, porque quien está en él ve contenido compartido por quien está en el tratamiento; por eso el experimento no cumple el supuesto conocido como SUTVA (que el resultado de una cuenta no dependa del tratamiento de las otras), lo que impide estimaciones no sesgadas de efectos causales, y las estadísticas probablemente los subestiman. Además, el tratamiento, es decir, la experiencia de ordenamiento, cambió a lo largo del tiempo. En cuanto al mecanismo, los autores dicen que entenderlo exige un estudio adicional y citan, como explicación posible, que partidos diferentes pueden seguir estrategias diferentes en Twitter. El grupo de control, por último, no se creó para la investigación, sino para mejorar el algoritmo y dar una línea de base al negocio; el trabajo fue revisado por los equipos jurídico y de privacidad de Twitter, y no por un comité de ética.

Mosleh y Rand (2022), descritos en el Capítulo 1, tocan el mismo problema desde el lado de la audiencia. Construyeron puntuaciones de falsedad para 816 élites públicas de los Estados Unidos a partir de las verificaciones de PolitiFact: cada verificación recibe una nota de veracidad (de 1, verdadero, a 0, “pants on fire”), y el promedio de las notas de la figura pública, restado de 1, da la puntuación de falsedad. Construyeron también, para cada usuario de Twitter, una puntuación de exposición basada en a quién sigue. La exposición a la desinformación de las élites se correlacionó negativamente con la calidad de las noticias que el propio usuario comparte y positivamente con la ideología conservadora estimada. Los autores dicen que el enfoque se basa en la red de cuentas seguidas y no capta la exposición más allá de ellas, como por retuits o por recomendaciones algorítmicas, y sugieren que la puntuación puede servir, en trabajos futuros, para distinguir el papel de las recomendaciones algorítmicas y el de las preferencias individuales. Las puntuaciones de falsedad vienen de un único verificador y dependen de lo que este verificó, una cuestión de selección que trató el Capítulo 2.

Lo que mide la auditoría es la diferencia entre dos programadores, el ordenamiento personalizado y el orden cronológico, y no lo que el ordenamiento optimiza ni si las reacciones entran en el cálculo. Para el argumento del capítulo, eso basta para una conclusión limitada: la plataforma es una variable que altera el alcance de cada emisión, y la alteración no es uniforme entre emisores, de modo que el análisis del productor necesita dos variables, lo que la audiencia hace con lo que recibe y cuánto recibe, y la segunda la fija el programador. Si la diferencia entre grupos refleja una preferencia de la plataforma, una es-

trategia de los partidos o la conducta de la audiencia es lo que los datos no deciden; se usa el mismo criterio para todos los campos: la amplificación es un dato de alcance, y no de sesgo ni de intención.

DATO	En seis de siete países, el mayor partido de derecha o centroderecha tuvo una amplificación por la personalización significativamente mayor que el mayor de izquierda o centroizquierda; en Alemania la diferencia no fue significativa; por político, no hubo asociación entre partido y amplificación; el resultado para las noticias depende de la base de etiquetas (Huszár et al., 2022).
INTERPRETACIÓN	La plataforma es una variable que determina el alcance de cada emisión de modo desigual entre productores, y el alcance entra en lo que el productor puede recibir (Skinner, 1957). La diferencia entre los dos ordenamientos no revela lo que el ordenamiento optimiza.
HIPÓTESIS	Si las reacciones recibidas realimentan el ordenamiento, la plataforma cierra el lazo de la Figura 8.1: lo que recibió reacciones pasa a mostrarse a más gente, y el productor recibe más de lo mismo. Ninguna fuente del capítulo muestra lo que el ordenamiento usa como entrada.

8.4 Reforzamiento intermitente e interacción

La analogía más repetida sobre las redes sociales es la de la máquina tragamonedas: el usuario tira de la palanca, y el premio llega de vez en cuando. Esta sección dice de dónde viene el vocabulario de la analogía, lo que la literatura experimental describe para cada programa y lo que las fuentes verificadas sostienen sobre el feed, que es poco.

CONCEPTO EN USO

Reforzamiento intermitente (Skinner, 1957)

Reforzamiento intermitente es aquel en el que solo una parte de las respuestas, o de las ocasiones, va seguida del reforzador. Para Skinner (1957), el reforzamiento meramente intermitente, como el de las tiradas casuales de dados, basta para mantener la fuerza de una respuesta; para ello, remite al estudio experimental de la conducta no verbal. En la conducta verbal, la intermitencia es la regla: el efecto depende de la presencia y de la actividad de un organismo reforzador, el oyente, cuya conducta no es inevitable ni, en general, previsible. Los programas nombrados (de razón y de intervalo, fijos y variables), de los que la analogía de la tragamonedas toma el nombre, vienen de la literatura experimental que resume el recuadro siguiente.

El argumento de Skinner tiene dos consecuencias para el capítulo. La primera es que, para quien habla, la intermitencia es una condición normal, y no una peculiaridad de las redes: el oyente puede no estar, puede no reaccionar, y el efecto sobre quien habla es intermitente por construcción. La segunda es que lo que la plataforma cambia no es la presencia de la intermitencia, sino su forma: cuántos oyentes pueden reaccionar (sección 8.3) y la visibilidad del recuento para el productor y para los otros oyentes. Esta es la lectura aquí propuesta.

CONCEPTO EN USO

Programas de reforzamiento (Ferster y Skinner, 1957)

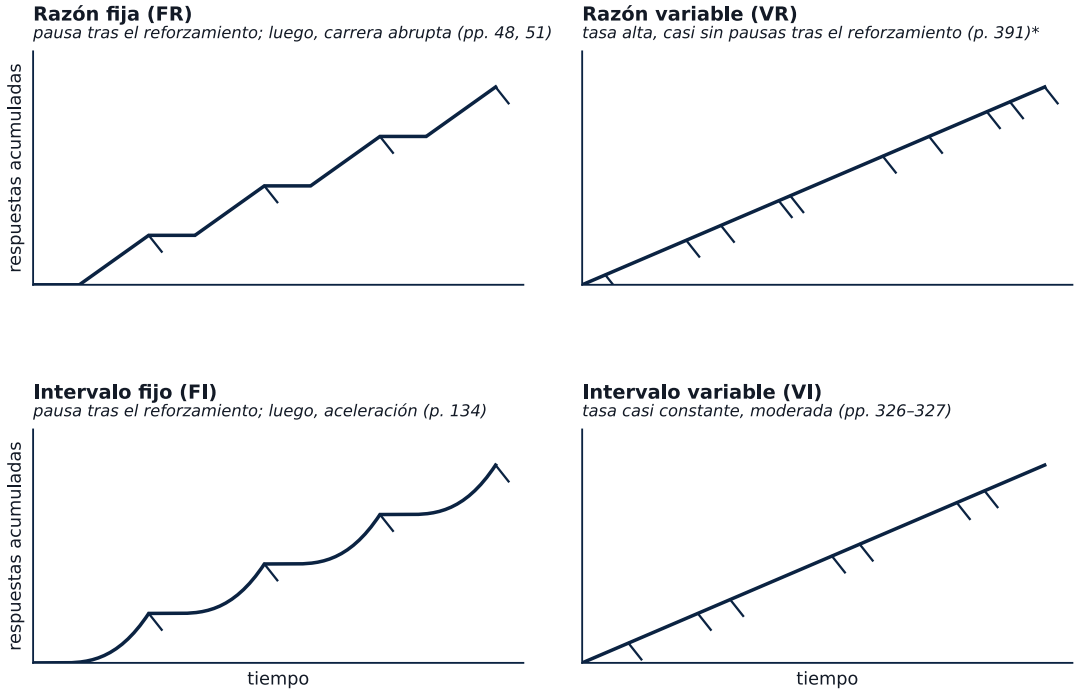
Un programa de reforzamiento puede definirse sin referencia a su efecto sobre la conducta: el reforzador puede depender del tiempo transcurrido desde el reforzamiento anterior o del número de respuestas emitidas desde entonces, y el programa puede ser fijo o variar, lo que da cuatro programas básicos. En la razón fija (FR), la respuesta se refuerza al completar un número fijo de respuestas contadas desde el reforzamiento anterior; en la razón variable (VR), según una serie aleatoria de razones con una media dada; en el intervalo fijo (FI), la primera respuesta después de un intervalo dado, medido desde el reforzamiento anterior; en el intervalo variable (VI), según una serie aleatoria de intervalos con una media dada. Los autores distinguen la función de moldear la forma de la respuesta y el control de estímulos de la función de mantener activo un repertorio, y dicen que la mayoría de los reforzamientos es intermitente. La mayor parte de los experimentos usa palomas que picotean una tecla para obtener alimento, y algunos usan ratas que presionan una palanca. Para el nivel humano, dicen que el análisis de los programas resultó útil en el estudio de la conducta psicótica y en técnicas educativas en el aula, y que otras aplicaciones, en derecho y penología, religión, industria y comercio, ofrecen una promesa considerable.

Lo que genera cada programa, en los experimentos de los autores, es el término de comparación que la analogía necesita. En la razón fija estabilizada tras una larga exposición, las pausas después del reforzamiento son ocasionales y, cuando ocurren, el paso a la tasa terminal es abrupto (FR 60, después de más de 4.000 reforzamientos); con razones mayores las pausas crecen, hasta 10 minutos en FR 120. En el intervalo fijo, el estado estable normal tiene una pausa después de cada reforzamiento, seguida de una aceleración hasta una tasa terminal en general moderada. El intervalo variable está diseñado para producir una tasa constante, al impedir que algún aspecto de la conducta del ave adquiera propiedades discriminativas, y la mayoría de los VI incluye un intervalo cero o cercano a cero, para evitar la pausa después del reforzamiento; en un ave en VI 1 minuto, después de 16 sesiones, la tasa quedó aproximadamente constante, sin variación correlacionada con característica alguna del

programa. La razón variable, como la fija, dispone reforzamiento diferencial de tasas altas, y hay pocos casos de reforzamiento después de una pausa. La Figura 8.2 traza esos patrones en un esquema.

FIGURA 8.2

Patrones de respuesta que Ferster y Skinner (1957) describen para los cuatro programas básicos: esquema conceptual



Esquema original, sin datos: las inclinaciones y duraciones son ilustrativas, las escalas verticales son independientes (no compare alturas entre paneles) y las marcas oblicuas indican reforzamientos. Traza lo que Ferster y Skinner (1957) describen en texto, para palomas que picotean una tecla para obtener alimento.
** Con medias grandes, la razón variable genera variedad, con pausas largas y ráfagas (pp. 394–397).*

Nota. Esquema original del autor, sin datos, a partir de las descripciones en texto de Ferster y Skinner (1957), cuyos registros son de palomas. El registro acumulativo, en el que las respuestas se suman a lo largo del tiempo, es el dato básico de los autores. Las inclinaciones y duraciones son ilustrativas, y la escala vertical de cada panel es independiente.

La razón variable, el programa que invoca la analogía, es también aquel en el que los autores registran mayor variedad: produce desempeños diversos según la distribución de los números de respuestas exigidas, y el ave deja de responder si la media pasa de cierto valor. Con una media de 360 respuestas, un ave mantuvo en una sesión una tasa sostenida de cerca de 7 respuestas por segundo y, en la siguiente, una ráfaga justo después del reforzamiento y luego pausas largas, una de más de 1 hora; los autores tratan los dos desempeños como los límites de un estado estable. La otra ave nunca sostuvo una tasa alta y respondía en ráfagas de 10 a 75 respuestas, separadas por pausas de 10 a 70 segundos.

Gunnarsson et al. (2026) escriben, en la discusión, que sus hallazgos ilustran cómo las plataformas emplean programas de razón variable: como los "me gusta" no están garantizados, cada episodio de atención exitoso se vuelve desproporcionadamente reforzante. Es interpretación de los autores; el programa no se manipuló ni se midió, y no podría haberse hecho en una tarea hipotética de compra. El argumento también salta de la intermitencia a la razón variable: que los "me gusta" no estén garantizados muestra que el reforzamiento es intermitente, y llamarlo razón variable exige, además, que la entrega dependa del número de respuestas y varíe en torno a una media (Ferster y Skinner, 1957). Los autores sugieren también que quien usa redes principalmente por reforzamiento ligado a la atención puede ser particularmente vulnerable a patrones de uso problemático, a partir de la interacción entre la escala de uso de redes y la condición de función para el consumo a costo mínimo.

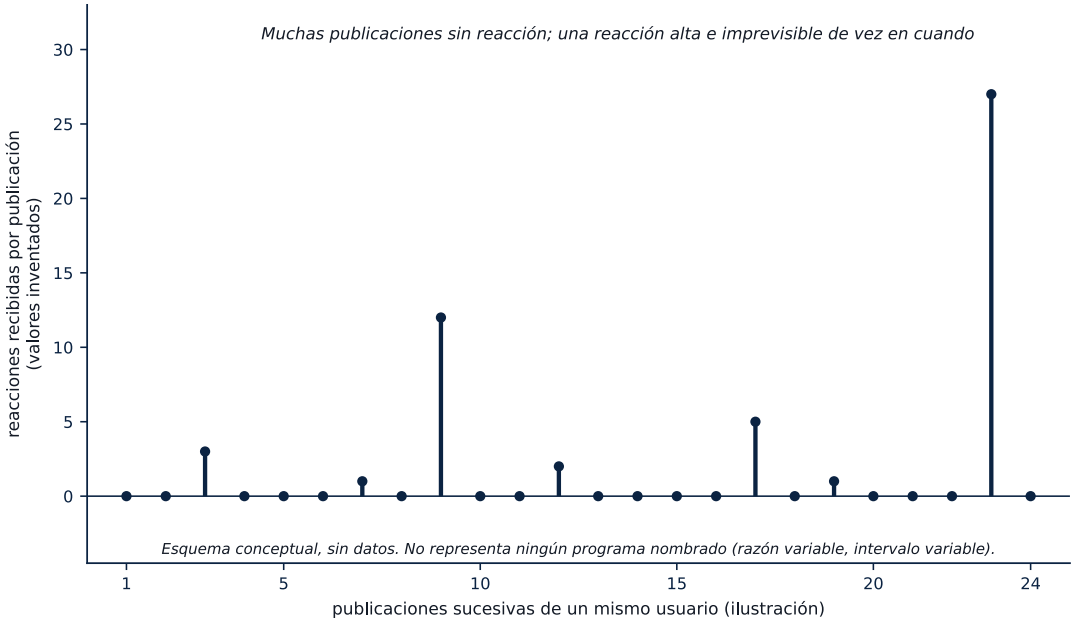
Dos estudios del Capítulo 9 son compatibles con la hipótesis sin ponerla a prueba. Lindström et al. (2021) muestran, en datos de Instagram y de foros, que la tasa de publicar sigue la tasa de "me gusta" según una función hiperbólica, que sube deprisa y después se aplana (compatible con la ley cuantitativa del efecto), mejor que según una relación lineal. Ceylan et al. (2023) recompensaron con puntos, en 80 ensayos, la elección de compartir o no titulares según un criterio de veracidad y muestran que el patrón de compartir, instalado por ese retorno, persiste en la prueba en la que el retorno se retira. En ninguno de los dos la intermitencia fue la variable manipulada: la persistencia sin retorno es compatible con una historia de reforzamiento intermitente y con otras historias, y el diseño no las separa. Lindström et al. tratan de cuándo se publica, y no de lo que se publica, y advierten que hay muchos motivos para publicar además de la recompensa (sección 9.2).

Llamar al feed programa de razón variable exige, como mínimo, tres cosas. La primera es la distribución de las reacciones por publicación, para un mismo productor, a lo largo de muchas publicaciones: la razón variable supone una serie de razones con una media dada, y no solo un recuento que oscila. La segunda es la relación entre el número de respuestas, publicar o revisar las reacciones, y la entrega del reforzador: en la razón, lo que cuenta es el número de respuestas desde el reforzamiento anterior, y en el intervalo, el tiempo. La tercera es el patrón de respuesta que genera el programa, comparado con el de la Figura 8.2 y con una línea de base, por ejemplo el tiempo hasta la publicación siguiente después de una con muchas reacciones y después de una ignorada. Ninguna fuente del capítulo trae la primera ni la segunda, y la tercera existe, para el feed, solo como pregunta. La Figura 8.3 sirve para ver los puntos flojos de la analogía, con valores inventados.

En la figura aparecen tres puntos flojos. El primero es que la reacción a una publicación no es un reforzador discreto que llega o no llega, como la ficha de una máquina o la operación del comedero en los experimentos de los autores: es un recuento, que crece con el tiempo y depende de cuántos ven (sección 8.3). El segundo es que hay al menos dos operantes candidatos, publicar y revisar las reacciones, y es posible que estén bajo reglas de entrega diferentes: el reforzador de revisar está disponible siempre que alguien reaccione, y el de publicar depende de la publicación. Los autores estudiaron, en experimentos que describen como exploratorios, dos respuestas con programas propios y simultáneos, y la de razón dominó a la de intervalo; que algo parecido ocurra entre publicar y revisar también es una pregunta, y no un resultado. El tercero es el más importante para este análisis: si lo que determina cuánto recibe cada publicación es, en parte, su forma, y no el azar, la contingencia es diferencial, pues selecciona formas y no solo dosifica la frecuencia, lo que sugeriría por lo menos que habría programas diferentes superpuestos o concurrentes entre sí. Un programa de reforzamiento describe el cuánto y el cuándo; la selección de formas describe el qué, y los autores distinguen el moldeamiento de la forma de la respuesta del mantenimiento del repertorio. La selección de formas es la que importa para la pregunta de este capítulo, y la que mejor sostienen las fuentes del capítulo, en la sección 8.2. Todo esto es interpretación.

FIGURA 8.3

Reacciones recibidas por publicaciones sucesivas de un mismo usuario: esquema conceptual



Nota. Esquema original del autor. Los valores se inventaron para mostrar la idea de reforzamiento intermitente y no representan datos, ni un programa nombrado (razón variable, intervalo variable).

DATO	Para Skinner (1957), la intermitencia es la regla de la conducta verbal y basta para mantener una respuesta; Ferster y Skinner (1957) definen los cuatro programas básicos y describen, en palomas, un patrón para cada uno (pausa y carrera en la razón fija, pausa y aceleración en el intervalo fijo, tasa casi constante en el intervalo variable, tasa alta con pausas raras en la razón variable, con variedad cuando la media es grande); Gunnarsson et al. (2026) atribuyen a las plataformas programas de razón variable sin medirlos; la tasa de publicar sigue la de "me gusta" según una función hiperbólica (Lindström et al., 2021).
INTERPRETACIÓN	La intermitencia es una condición general de la conducta verbal. Que la entrega de reacciones a quien publica sea intermitente es plausible y coherente con los datos, pero es una premisa, y no un resultado; y la intermitencia no es, por sí misma, razón variable.
HIPÓTESIS	La entrega sigue un programa nombrado (razón variable o intervalo variable) y el patrón de uso se deriva de ello. Exigiría la distribución de reacciones por publicación, la relación entre respuestas y entrega y el patrón de respuesta comparado con el de la Figura 8.2; ninguna fuente del capítulo trae los dos primeros, y el tercero no se ha examinado para el feed.

8.5 Experimentos con el propio algoritmo: lo que cambió y lo que no cambió

Los experimentos en los que se manipuló la regla de entrega son la evidencia más directa de lo que hace la plataforma. Las fuentes verificadas traen dos, el de Huszár et al. (2022), descrito en la sección 8.3, y el de Nyhan et al. (2023), además de la prueba del editor de Robertson et al. (2023), que manipula otro punto de la cadena. La Tabla 8.2 los pone lado a lado.

El experimento de Nyhan et al. (2023) está descrito en la sección 6.2, y aquí interesa lo que manipuló. Para los participantes asignados al tratamiento, todo el contenido de amigos, grupos y páginas del que se predecía que compartía la inclinación política del participante fue relegado en el feed, del 24 de septiembre al 23 de diciembre de 2020; el del grupo de control no se alteró de modo sistemático. Según los autores, el único objetivo de la intervención era reducir la exposición a fuentes de la misma inclinación, con la mayor intensidad de relegamiento que, según indicó una prueba previa, no dejaría el feed casi vacío. La exposición cayó del 53,7% al 36,2%, y no hubo un efecto medible

en las ocho medidas de actitud prerregistradas; quienes tuvieron la exposición reducida interactuaron más, por exposición, con el contenido de la misma inclinación que aún veían.

TABLA 8.2

Tres manipulaciones a lo largo de la cadena productor–plataforma–audiencia: lo que cambió y lo que no cambió

Estudio	Lo que se manipuló	Lo que midieron	Lo que cambió	Lo que no cambió o no se midió
Robertson et al. (2023)	Variaciones de titular de la misma noticia, asignadas al azar a lectores de Upworthy (manipulación del editor, no de la plataforma)	Tasa de clics de cada variación	Las palabras negativas aumentaron los clics (2,3% por palabra, en un titular de longitud media); las palabras positivas los redujeron (cerca del 1,0%)	Lo que el editor pasó a producir; el individuo; la experiencia del lector
Huszár et al. (2022)	Orden del feed: ordenamiento personalizado (tratamiento, 4% de las cuentas) frente a orden cronológico (control, 1%)	Alcance, en impresiones de permanencia, de tuits de legisladores de siete países y de enlaces de noticias	Alcance mayor con la personalización para todos los partidos; mayor para el mayor partido de derecha o centroderecha en seis de los siete países	Asociación entre partido y amplificación, por político; mecanismo causal; efectos indirectos sobre el control; lo que los productores hicieron después
Nyhan et al. (2023)	Relegamiento, en el feed de Facebook, del contenido de fuentes de la misma inclinación (24/09 a 23/12/2020; n = 23.377)	Proporción de contenido de la misma inclinación y de fuentes transversales; ocho medidas de actitud prerregistradas	Exposición a fuentes de la misma inclinación del 53,7% al 36,2%, y a fuentes transversales del 20,7% al 27,9%; más interacción, por exposición, con el contenido afín restante	Actitudes (efectos de ±0,12 desviaciones estándar o más descartados); efectos del uso previo y acumulados en años; el productor

Nota. Los números vienen de los artículos. Las páginas de Robertson et al. (2023) son las de Nature Human Behaviour y las de Nyhan et al. (2023), las de Nature; Huszár et al. (2022) se leyó sin paginación impresa. “Cambió” y “no cambió” se refieren a lo que midió cada estudio.

La sección 6.1 dejó una pregunta sobre la frontera de la burbuja: cuánto de lo que viene de fuera de la comunidad verbal llega al hablante. La regla de entrega es una de las variables que lo deciden, y el experimento muestra que la mueve: con el relegamiento del contenido afín, la exposición a fuentes transversales subió del 20,7% al 27,9%, y la de fuentes ni afines ni transversales, del 25,6% al 35,9%. Lo que el experimento no mide es lo que preguntaba la figura de la sección 6.1: si la corrección dirigida a lo que dijo el hablante pasó a llegarle. La exposición a fuentes transversales es una proporción del feed, y no una corrección del enunciado.

Lo que los dos experimentos tienen en común es lo que no miden. Ambos intervienen en la entrega y miden lo que el oyente recibe o declara: el alcance de tuits de políticos, en un caso; la exposición y las actitudes de usuarios, en el otro. Ninguno mide lo que el productor hizo después de ver cambiar su alcance. El ordenamiento de Huszár et al. alteró el tamaño de la audiencia de cada emisión, y el relegamiento de Nyhan et al. redujo el alcance del contenido de la misma inclinación; para el productor, ambos son cambios en la contingencia de alcance. La pregunta de la que partió el capítulo, si quien publica ajusta lo que publica a esa regla y cómo, solo se respondería con series de publicaciones de productores antes y después de cambios en la regla de entrega. Las plataformas tienen esos registros; los estudios verificados no los traen, lo que limita el acceso a las posibles funciones atencionales diferenciales implicadas.

Nada de esto autoriza a concluir que la regla de entrega no importa (Nyhan et al. no encontraron efecto en las actitudes declaradas, después de tres meses y en un único servicio) ni que explique lo que se dice (Huszár et al. midieron alcance, y no lo que se dice). El efecto nulo en las actitudes tiene, por cierto, explicaciones que los propios autores enumeran: el contenido político afín es una fracción de lo que se ve, Facebook es una fracción de la dieta informativa, persuadir es difícil y el diseño no capta el uso previo ni los efectos acumulados a lo largo de los años.

DATO	El ordenamiento personalizado y el orden cronológico difieren en el alcance de partidos y de fuentes de noticias (Huszár et al., 2022); reducir en un tercio la exposición a contenido de la misma inclinación no alteró ocho medidas de actitud en tres meses (Nyhan et al., 2023).
INTERPRETACIÓN	La regla de entrega altera el alcance, y el alcance es una variable de la que depende lo que el productor puede recibir. En los dos experimentos la variable dependiente es del oyente o del alcance, y no del productor.

HIPÓTESIS Los productores ajustan lo que publican a los cambios en el alcance que produce la regla de entrega. Para ponerla a prueba, habría que medir la producción antes y después de un cambio de regla; ninguna fuente verificada lo hace.

8.6 Interpretaciones rivales y límites

La lectura del capítulo, según la cual la atención distribuida por el algoritmo funciona como reforzador de quien publica, tiene rivales. El recuadro de abajo las reúne, una por una, y los límites del capítulo vienen después de él.

Interpretaciones rivales para el mismo conjunto de datos

El usuario elige, y el algoritmo refleja. Aquello a lo que la audiencia reacciona y lo que el algoritmo muestra se confunden: Mosleh y Rand (2022) dicen que su puntuación no capta la exposición por recomendación algorítmica y sugieren que sirva para separar el papel de las recomendaciones y el de las preferencias; Nyhan et al. (2023) señalan que los cambios probados en el algoritmo no anulan la propensión de los usuarios a buscar información afín; y Rathje et al. (2021) dicen no saber cómo contribuyen las elecciones algorítmicas.

La atención no es el único reforzador, ni el principal. En la tarea de Gunnarsson et al. (2026), el escape, la información y el reforzamiento positivo automático generaron más demanda que buscar "me gusta", y los autores recuerdan que muchas personas usan las redes por funciones múltiples y superpuestas.

Norma e identidad en lugar de reforzamiento. Lo que se observa puede ser conformidad con las normas de la red (Brady et al., 2021) o expresión de identidad de grupo (Rathje et al., 2021); véase la sección 9.8.

Novedad y clickbait. El efecto de las palabras negativas puede ser específico de un medio de clickbait y de una fase en la que el recurso era nuevo para los lectores (Robertson et al., 2023).

Efectos pequeños. Los efectos son del 2 al 3% (Brady et al., 2021) y de cerca del 2,3% por palabra (Robertson et al., 2023); pueden escalar, pero eso es un argumento, y no un resultado.

Alcance no es sesgo ni intención. La amplificación por partido o por fuente es un dato de alcance, depende de la base de etiquetas y no tiene un mecanismo identificado (Huszár et al., 2022).

Datos que solo tiene la plataforma. El grupo de control de Twitter existía para el negocio y fue revisado por los equipos jurídico y de privacidad, y no por un comité de ética; el tratamiento cambió a lo largo del tiempo (Huszár et al., 2022). Quien, desde fuera, no tiene acceso a los registros no replica ni amplía el estudio.

Cuatro límites valen para todo el capítulo. Primero, ninguna fuente es brasileña, y las de plataforma tratan de Twitter y de Facebook, en los Estados Unidos y, en el caso de Huszár et al. (2022), en otros seis países entre los que no está Brasil; las plataformas cambian (el tratamiento del experimento de Twitter cambió a lo largo del tiempo), y nada de lo que aquí se dice trata de WhatsApp, YouTube o TikTok. Segundo, el dato de mayor alcance sobre la regla de entrega (Huszár et al., 2022) mide alcance y no conducta del productor. Tercero, los datos sobre el contenido moralizado son, en su mayor parte, de asociación, y los efectos son pequeños. Cuarto, el capítulo no muestra que el feed sea un programa de reforzamiento: muestra que, para la conducta verbal, la intermitencia es la regla y que la regla de entrega de la plataforma altera el alcance; el resto es hipótesis que pide datos que solo tienen las plataformas.

Para quien quiera usar el capítulo en el análisis de declaraciones (Capítulos 3, 5 y 7), la consecuencia es práctica. El alcance de una declaración es una variable que el analista de la conducta no controla y que altera lo que ella puede recibir; comparar la declaración de un político con la de otro sin registrar el alcance de cada una, y lo que la plataforma hizo con él, mezcla dos variables. Cuando el alcance no se conoce, el analista debe decirlo.

Síntesis

La atención del oyente es, en el análisis de Skinner (1957), un reforzador generalizado que fortalece la conducta verbal con independencia de la forma, y en las plataformas lo que llega al productor es un recuento de reacciones que se aproxima a ella. La hipótesis de que ese recuento sea el reforzador dominante de quien publica tiene apoyo parcial en los estudios del Capítulo 9 y una cautela en el de Gunnarsson et al. (2026), en el que la demanda de usar las redes para obtener atención fue baja (sección 8.1). Los datos sobre el contenido emocional y moralizado no forman un solo hallazgo: las palabras negativas aumentan los clics, la referencia al exogrupo aumenta las veces compartido, y la indignación sigue el retorno recibido; son respuestas distintas, medidas por propiedades de forma, y la función que las mantiene no se midió (8.2).

La plataforma es el tercer término. El ordenamiento fija el alcance de cada emisión, y la auditoría de Twitter muestra que difiere del orden cronológico, con una amplificación mayor para el mayor partido de derecha o centroderecha en seis de siete países, ninguna asociación por político y resultados para las noticias que dependen de la base de etiquetas; el mecanismo es desconocido (8.3). La hipótesis de que la entrega siga un programa de razón variable es coherente con Skinner, que da la intermitencia como regla de la conducta verbal, pero ninguna fuente la mide; la literatura sobre los programas (Ferster y

Skinner, 1957) da, para palomas, un patrón por programa, con variedad en la razón variable de medias grandes, y la comparación con el feed pide datos que solo tienen las plataformas. La analogía tiene además puntos flojos: la reacción es un recuento, hay dos operantes candidatos y, si la forma de la publicación determina lo que recibe, lo que hay es selección de formas (8.4). Los dos experimentos con el algoritmo cambiaron el alcance y la exposición y no cambiaron las actitudes declaradas; ninguno midió lo que el productor hizo después (8.5).

Los Capítulos 3, 5 y 7 toman de este capítulo la idea de que el alcance es una variable que hay que registrar en el análisis de cualquier declaración; el Capítulo 3, además, la distinción entre la función y la forma que las propiedades del contenido dejan abierta. El Capítulo 6 recibe una respuesta parcial a la pregunta de cuánto vuelve permeable la regla de entrega la frontera de la burbuja: la mueve, y lo que atraviesa la frontera como corrección no se midió. El Capítulo 11 (sección 11.2), al tratar de las medidas de difusión, retoma la distinción entre alcance, exposición y reacción. El Capítulo 12 (sección 12.6) parte de lo que cambió y de lo que no cambió con la regla de entrega para discutir intervenciones sobre las contingencias de la plataforma.

Referencias

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), Artículo 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), Artículo 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*. Publicación anticipada en línea. 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), Artículo 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), Artículo 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), Artículo 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), Artículo 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.

instituto WALDEN4 | ACADEMIA WALDEN4

MATERIALES EDUCATIVOS

Por dónde empezar a leer Análisis de la Conducta

Artículos científicos, revistas e instituciones de referencia reunidos para la formación en Análisis de la Conducta y en práctica basada en la evidencia.

Acceda, gratis y sin registro

walden4.com.br/materiais

CAPÍTULO

9

El ciclo productor–audiencia

Hildum y Brown publicaron en 1956 un experimento en el que un entrevistador telefoneó a 40 estudiantes varones de la escuela de verano de Harvard y les leyó 15 afirmaciones sobre la filosofía de Educación General de la universidad, cada una con cuatro alternativas, de "totalmente de acuerdo" a "totalmente en desacuerdo". La mitad de ellos oyó "Good" ("bien"), en tono neutro, después de algunas respuestas; la otra mitad oyó "Mm-hmm" en los mismos puntos. En cada mitad, diez estudiantes oían la reacción después de las respuestas favorables a la Educación General y diez, después de las desfavorables (Hildum & Brown, 1956). Con el "Good", la media de actitud (menor es más favorable) fue de 24,95 en el grupo que recibió la reacción en las respuestas favorables y de 31,75 en el que la recibió en las desfavorables. De los 20 estudiantes que oyeron "Good", ocho notaron que el entrevistador decía la palabra, y solo uno supuso que era aprobación; todos negaron que sus respuestas hubieran sido influidas por él.

La pregunta de los autores era de método: si un entrevistador puede influir en el entrevistado sin querer, la influencia opera fuera de la conciencia de todos los implicados y la encuesta de opinión queda sesgada. Este capítulo hace la pregunta desde el lado de quien habla. ¿Qué hace la reacción de quien oye con lo que se dice, y qué hace lo que se dice con quien oye? En 1956 la reacción era una palabra de un entrevistador. Hoy es un recuento público de "me gusta" y de veces compartido, distribuido por un algoritmo. El capítulo reúne primero la evidencia experimental más antigua (sección 9.1) y después la de plataformas (9.2, 9.3, 9.5 y 9.6), propone un esquema que las une (9.4) y una hipótesis de nivel cultural (9.7), y dice en cada paso qué es dato, qué es interpretación y qué es hipótesis.

9.1 Las opiniones pueden reforzarse: los experimentos de 1956 a 1969

El experimento de Hildum y Brown, que abre el capítulo, tiene antecesores y sucesores. Los estudios de reforzamiento verbal anteriores a él tomaban como unidad de respuesta una palabra o una clase de palabras. En uno de ellos, el experimentador murmuraba "Mm-hmm" cuando el estudiante decía un sustantivo en plural; en otros, decía "Good" después de frases que empezaban con el pronombre elegido; en todos, los sujetos decían no haber percibido relación entre lo que decían y la reacción (Hildum & Brown, 1956). Los autores argumentan que un entrevistador de opinión reaccionaría a categorías de contenido, y no a palabras específicas, y usaron el teléfono para que la reacción vocal fuera la única posible.

Scott (1957) creó la situación de aprobación en el aula. Estudiantes de 29 grupos de psicología general escribieron lo que pensaban sobre tres cuestiones polémicas de la universidad (servicio militar universal, horario nocturno de las universitarias y énfasis en el fútbol), y cada grupo recibió una de ellas para debatir. En cada grupo, el instructor eligió a dos personas claramente favorables y a dos claramente contrarias y las invitó a un "experimento" sobre cuánto lograrían influir en el grupo, con la condición de que cada una defendiera el lado opuesto al propio. Dos semanas después, los pares debatieron ante el grupo, con tres minutos de exposición y dos de réplica para cada lado, y el grupo votó en papeletas qué participante del debate lo había hecho mejor. La votación era fingida: el instructor simulaba contar las papeletas, y quién "ganaría" en cada par se había definido de antemano, alternando el lado ganador. Justo después, todos volvieron a escribir su opinión sobre la cuestión, y los textos, copiados en tarjetas identificadas solo por número, se codificaron en una escala de siete puntos, de muy favorable a muy contrario. De los 58 pares designados, 36 produjeron datos completos, en 19 grupos y con seis instructores; otros 36 estudiantes, con una opinión clara pero sin debate, formaron el control. Los ganadores cambiaron en promedio 1,25 puntos en la dirección del lado que habían defendido, los perdedores cambiaron 0,17 puntos en sentido contrario, y los controles, 0,31 puntos en el mismo sentido que los ganadores. La diferencia entre ganadores y perdedores fue estadísticamente significativa ($t = 3{,}97$; $p < 0{,}001$); la t es el estadístico de la prueba que compara dos medias, y la p es la probabilidad de obtener una diferencia tan grande, o mayor, si en realidad los grupos no difirieran, de modo que, por convención, una p por debajo de 0,05 se llama "significativa". La diferencia entre ganadores y controles también fue significativa ($p < 0{,}02$), y la diferencia entre perdedores y controles no (Scott, 1957).

Spielberger et al. (1962) estudiaron el reforzamiento verbal en una tarea que no es de opinión, la construcción de frases, y preguntaron si el efecto exige que la persona perciba la contingencia. Cada una de 45 universitarias, evaluada individualmente, recibía 100 tarjetas, con un verbo en pasado y los pronombres *I*, *we*, *you*, *he*, *she* y *they* en órdenes diferentes, y debía componer una frase que empezara por uno de los pronombres y contuviera el verbo; cada frase era un ensayo. En los 20 primeros no se reforzó a nadie, lo que mide la frecuencia de base de frases iniciadas por *I* o *we*. Después, las participantes fueron sorteadas en tres grupos de 15, que recibieron instrucciones neutras, instrucciones que pedían la primera frase que viniera a la mente, con prisa, y desalentaban la búsqueda de un principio, o instrucciones que decían que existía un modo correcto de armar las frases, para que la percepción de la contingencia variara. En los 80 ensayos siguientes, en cada grupo 10 oyeron "good", dicho en tono neutro, después de cada frase iniciada por *I* o *we*, y cinco, que formaban el control, no oyeron nada. Al final, una entrevista de preguntas cada vez más específicas, que evitaba sugerir la contingencia correcta, clasificó a cada participante como consciente o no de que el "good" seguía a aquellas frases. De las 30 reforzadas, 19 fueron clasificadas como conscientes; solo ellas aumentaron el uso de *I* y *we*, y ninguna de las 11 no conscientes aumentó en más de una respuesta (Spielberger et al., 1962).

Insko (1965) llevó el procedimiento al teléfono y midió si el efecto duraba. Seis asistentes de posgrado telefonearon, cada uno, a 12 estudiantes sorteados del curso de introducción a la psicología de la Universidad de Hawái (72 en total) y les leyeron 14 afirmaciones de opinión sobre la creación de una Semana Aloha en primavera, como la de que traería más dinero de turistas o más congestión; las afirmaciones no decían que la idea fuera buena o mala. El estudiante respondía si estaba de acuerdo o en desacuerdo con cada una, y cada asistente dijo "good" después de las respuestas de actitud positiva para seis de los estudiantes y después de las de actitud negativa para los otros seis. Cada respuesta valía de 0 a 3 puntos, y la suma media fue de 28,22 entre los reforzados en las respuestas positivas y de 20,69 entre los reforzados en las negativas. Cerca de una semana después, el experimentador, que era el profesor del grupo, distribuyó a los 225 alumnos un cuestionario sobre asuntos locales, con el ítem sobre la Semana Aloha hacia los dos tercios del cuestionario. De los 72 contactados por teléfono, 70 respondieron. El grupo antes reforzado en las respuestas positivas quedó, en promedio, en 10,40, en la región de la escala rotulada "algo a favor", y el antes reforzado en las negativas, en 8,20, en la región "indeciso" ($t = 2{,}03$; $p < 0{,}05$) (Insko, 1965).

Insko y Melson (1969) usaron el mismo procedimiento para preguntar si el efecto depende de que la persona sepa que está en un experimento, la hipótesis de las características de demanda: el participante actuaría como supone que el experimentador espera. En el experimento I, 72 estudiantes varones del curso de introducción a la psicología de la Universidad de Carolina del Norte fueron contactados, la mitad por el procedimiento habitual de reclutamiento para experimentos de laboratorio y la mitad por teléfono. En ambos casos, el entrevistador se presentaba como miembro de un centro de investigación por encuestas que sondeaba la actitud de los estudiantes sobre la televisión de pago, y a quien recibió la llamada no se le informó de que participaba en un experimento. El participante estaba de acuerdo o en desacuerdo con 12 afirmaciones de opinión sobre la televisión de pago, y el "good" seguía a estar de acuerdo con afirmaciones a favor y en desacuerdo con afirmaciones en contra (condición pro) o al revés (condición contra); el momento del semestre, inicio o fin, también se varió. Justo después, todos respondieron a un cuestionario sobre la percepción del reforzamiento, elaborado por Dulany en 1962. El experimento II repitió el diseño con 20 sujetos en cada una de las cuatro combinaciones de contexto (laboratorio o teléfono) y dirección del reforzamiento, y añadió una línea de base de ocho afirmaciones sin reforzamiento. En los dos, el único efecto significativo fue el del reforzamiento (F = 6,10 y 5,62; la F es el estadístico del análisis de varianza, que compara la variación entre los grupos con la variación dentro de ellos), sin diferencia entre laboratorio y teléfono (Insko & Melson, 1969).

La Tabla 9.1 reúne los cinco estudios, con lo que se hizo, lo que se observó y lo que los resultados no permiten decir. Cuatro tratan de opiniones y actitudes; el de Spielberger et al. usa una tarea que no es de opinión y entra por ser la principal evidencia contraria sobre la percepción de la contingencia.

TABLA 9.1

Experimentos de reforzamiento verbal de opiniones y actitudes, de 1956 a 1969: lo que se hizo, lo que se observó y lo que no permiten decir

Estudio	Participantes y tarea	Reacción a lo que se dijo	Lo que se observó	Salvedad
Hildum y Brown (1956)	40 estudiantes varones (escuela de verano de Harvard), 10 por grupo; 15 afirmaciones sobre la Educación General, por teléfono	"Good" o "Mm-hmm" después de respuestas favorables o de desfavorables	Con "Good", media de actitud de 24,95 (reacción a las favorables) frente a 31,75 (a las desfavorables; t = 3,31; p = 0,01); con "Mm-hmm", diferencia no significativa y en sentido opuesto; replicación con 25 estudiantes, mismo patrón	8 de 20 notaron el "Good"; todos negaron influencia; de los 24 que aventuraron una conjetura, todos menos uno dijeron que el entrevistador estaba a favor de la Educación General
Scott (1957)	72 participantes en debates (36 pares) de 19 grupos de psicología, que defendían la posición opuesta a la propia; 36 controles	Voto del grupo, definido de antemano, que declaraba ganador o perdedor	Los ganadores cambiaron en promedio 1,25 puntos (escala de 7) en la dirección del debate; los perdedores, 0,17 en sentido opuesto; los controles, 0,31 (ganadores × perdedores: t = 3,97; p < 0,001)	Sin condición de debate sin voto; postest en un máximo de cinco minutos; solo siete ganadores invirtieron su opinión; persistencia no evaluada
Spielberger et al. (1962)	45 universitarias; construcción de frases con pronombres (tarea que no es de opinión); 30 reforzadas y 15 controles	"Good" después de frases iniciadas por I o we	Solo los 19 sujetos que verbalizaron la contingencia aumentaron el responder; ninguno de los 11 sin percepción subió más de una respuesta	Los autores advierten sobre el papel de la tarea y de la población; la actitud ante el reforzamiento moduló el efecto entre los conscientes
Insko (1965)	72 estudiantes (Hawái); 14 afirmaciones sobre una Semana Aloha de primavera, por teléfono; cuestionario en el aula cerca de una semana después (N = 70)	"Good" después de estar de acuerdo o después de estar en desacuerdo	Teléfono: 28,22 (reacción a la actitud positiva) frente a 20,69 (a la negativa); cuestionario: 10,40 frente a 8,20 (t = 2,03; p < 0,05)	El autor habla de generalidad de respuesta; el experimentador era el profesor del grupo que respondió al cuestionario

Estudio	Participantes y tarea	Reacción a lo que se dijo	Lo que se observó	Salvedad
Insko y Melson (1969)	Exp. I: 72 varones; Exp. II: 80 (20 por celda); teléfono o laboratorio; 12 afirmaciones sobre la televisión de pago	"Good" después de respuestas pro o contra	Efecto del reforzamiento significativo (F = 6,10 y 5,62) e igual en el laboratorio y por teléfono; entre quienes no percibieron la contingencia, diferencia no significativa en cada experimento; combinados, 6 de 6 comparaciones en la dirección prevista (p ≈ 0,016)	Efecto menos robusto que en un estudio anterior con las mismas afirmaciones (F = 23,09); solo 14 de 152 informaron de alguna intención

Nota. Los números vienen de los artículos (Hildum & Brown, 1956; Scott, 1957; Spielberger et al., 1962; Insko, 1965; Insko & Melson, 1969). En Hildum y Brown, una puntuación menor es una actitud más favorable; en Insko, una mayor. "Pro" y "contra" se refieren al tema de la entrevista.

La tabla muestra un patrón con cuatro partes. Primera: en las tareas de opinión, la aprobación de quien oía desplazó lo que se declaró en la dirección reforzada, en los cuatro estudios. Segunda: el efecto apareció fuera de la situación en la que se produjo. Insko (1965) lo encontró una semana después, en otro medio (cuestionario escrito) y en otro contexto (de la casa al aula), e Insko y Melson (1969) lo encontraron del mismo tamaño en el laboratorio y por teléfono, con sujetos que no sabían que estaban en un experimento. Tercera: el efecto es pequeño y frágil. En Scott (1957), solo siete de los ganadores invirtieron su opinión, los otros 14 que cambiaron solo la debilitaron, y el postest llegó como máximo cinco minutos después del voto; en Insko y Melson (1969), el efecto fue bastante menos robusto que en un estudio anterior con las mismas afirmaciones. Y el reforzador importa: el "Mm-hmm" no funcionó en Hildum y Brown, que sugieren que la diferencia puede venir de la mayor visibilidad del "Good" (p. 110) y, en parte, de su significado más claramente favorable. Cuarta: la percepción de la contingencia está en disputa. En Spielberger et al. (1962), solo aprendieron los sujetos que la verbalizaron; en Insko y Melson (1969), entre quienes no informaron percibir la contingencia la diferencia entre las condiciones no fue significativa en ninguno de los dos experimentos tomados por separado, pero las seis comparaciones combinadas fueron en la dirección prevista, y el nuevo análisis de un estudio anterior con las mismas afirmaciones dio una diferencia significativa. Los autores concluyen que el reforzamiento tuvo efecto también sin percepción y atribuyen la

diferencia con respecto a la construcción de frases a la tarea: opinar es una respuesta sobreaprendida, es decir, muy practicada, que se produce en pocos ensayos y con poca novedad.

CONCEPTO EN USO

Reforzamiento social generalizado y audiencia (Guerin, 1994), retomado del Capítulo 3

El Capítulo 3 (sección 3.3) presentó el reforzamiento social generalizado y la definición de Guerin (1994) de la actitud como enunciado mantenido por él. Para el ciclo, interesan dos consecuencias. La primera es que la consecuencia viene de otras personas y, en parte, no depende del contenido, como en el intercambio de actitudes que funciona como evento social ritual, en el que el reforzamiento está en el propio intercambio. La segunda es que la audiencia es estímulo discriminativo: audiencias específicas oyen y responden a actitudes específicas, y por eso lo que se declara tiende a atender a las contingencias de la comunidad que ocasionó la declaración. El reforzador generalizado no opera igual para todos. Spielberger et al. (1962) recuerdan que los estudios de condicionamiento verbal suponen que "Good" y "Mm-hmm" son reforzadores condicionados generalizados, con las mismas propiedades para todos los sujetos, y citan evidencia de que el "Mm-hmm" elevó el desempeño de quienes lo evaluaban como positivo y lo redujo en quienes lo evaluaban como negativo (Mandler y Kaplan, 1956, citados por Spielberger et al., 1962).

Interpretaciones rivales para el desplazamiento de la opinión bajo aprobación

Percepción de la contingencia. Si el hablante percibe que ciertas respuestas son aprobadas y pasa a darlas más, lo que se ve es un desempeño bajo una regla, y no un operante moldeado sin percepción. Los resultados se dividen según la tarea (Spielberger et al., 1962; Insko & Melson, 1969), y Hildum y Brown (1956) admiten que sus medidas de percepción pueden haber sido poco sensibles.

Características de demanda. Si el participante sabe que está en un experimento y actúa como supone que se espera, el efecto debería ser mayor en el laboratorio que fuera de él. Insko y Melson (1969) no encontraron diferencia entre el laboratorio y el teléfono sin identificación como experimento.

Respuesta de entrevista, y no actitud. El desplazamiento puede valer solo para lo que se dice en esa entrevista. Insko (1965) midió una semana después y en otro medio, y el propio autor describe lo que mostró como generalidad de respuesta.

Contacto con argumentos opuestos. En Scott (1957), ganadores y perdedores tuvieron el mismo contacto con los argumentos opuestos, y la diferencia entre ellos apunta a la aprobación o la desaprobación del grupo. Pero el cambio de los ganadores puede ser efecto solo del reforzamiento o efecto del contacto, que el reforzamiento mantuvo, y faltó una condición de participantes en debates sin voto para separar las dos lecturas.

Suposición sobre el entrevistador. En Hildum y Brown (1956), casi todos los que aventuraron una conjetura supusieron que el entrevistador estaba a favor de la Educación General, lo que los autores atribuyen a la suposición sobre quién investiga el tema, y no al patrón de reforzamiento.

DATO	En entrevistas por teléfono (Hildum & Brown, 1956; Insko, 1965; Insko & Melson, 1969) y en un debate en el aula (Scott, 1957), lo que las personas declararon como opinión se desplazó en la dirección de la aprobación recibida más que en la del grupo contrario o de control, con efectos pequeños y que varían con el reforzador y con la tarea.
INTERPRETACIÓN	La opinión declarada es sensible a las consecuencias sociales que la audiencia dispone para lo que se declara, como prevé el análisis de Guerin (1994). La reacción del oyente funciona como consecuencia del enunciado, y la audiencia, como ocasión para él.
HIPÓTESIS	La misma sensibilidad vale para lo que se publica cuando la reacción de la audiencia es un "me gusta" o un reenvío. Ninguno de los cinco estudios usa redes sociales ni reacciones públicas y contadas; la extrapolación se hace aquí.

9.2 Lo que recibe el productor: "me gusta", veces compartido y alcance

En las plataformas, quien habla es un productor: publica y recibe de vuelta, contadas y visibles, las reacciones del público, el retorno (*feedback*) de los estudios. Los estudios usados aquí son de dos tipos: observacionales, que preguntan si el acto de publicar sigue el retorno recibido, y experimentos en entornos simulados, en los que el retorno se manipula.

Lindström et al. (2021) trataron el acto de publicar como conducta operante libre, sin ensayos discretos: el equivalente de un animal ante una única palanca que puede presionarse en cualquier momento, con el "me gusta" como consecuencia. Usaron cuatro conjuntos de datos, que suman más de un millón de publicaciones de más de 4.000 usuarios: Instagram (estudio 1, con 2.039 usuarios y, en promedio, 418 publicaciones por persona) y tres foros temáticos, de moda masculina, de moda femenina y de jardinería (estudio 2, con 2.127 usuarios y, en promedio, 91 publicaciones por persona). Los foros entraron porque en Instagram hay motivos económicos para publicar y la posibilidad de cuentas fraudulentas y de "me gusta" falsos, que en los foros son menos probables.

Los autores hicieron dos análisis. En el primero, compararon dos formas de describir cómo varía la tasa de publicación con la tasa de "me gusta". La investigación con animales mostró que la tasa de respuestas suele crecer con la tasa de recompensas según una función hiperbólica, que sube deprisa y después se aplana, la llamada ley cuantitativa del efecto, considerada la firma de la conducta guiada por la recompensa. En los cuatro conjuntos, esa función describió los datos mejor que una relación lineal (R^2 medio de 0,43 en el estudio 1 y de 0,37 en el estudio 2; el R^2 es la fracción de la variación de los datos que el modelo explica). En el segundo, definieron la latencia entre dos publicaciones sucesivas (τPost) y pusieron a prueba si el historial de "me gusta" la predecía mediante la causalidad de Granger, que se establece cuando una variable, aquí los "me gusta", mejora la predicción de otra, aquí el intervalo, más allá de lo que ya predicen los valores anteriores de esta. Los "me gusta" anteriores predijeron el intervalo hasta la publicación siguiente en los cuatro conjuntos. Por último, con un modelo de aprendizaje, estimaron para cada usuario la tasa media de recompensa que venía recibiendo. En Instagram, el intervalo entre publicaciones fue cerca de un 18% menor, algo así como ocho horas, cuando esa tasa estimada era alta que cuando era baja.

El estudio 3 fue un experimento, con 176 participantes. Durante una sesión en línea de 25 minutos, podían publicar memes cuando quisieran y cuantas veces quisieran y recibían de 0 a 19 "me gusta" de supuestos otros usuarios. En el paso de la primera a la segunda mitad de la sesión, la tasa media de "me gusta" cambiaba, de 0 a 9 por publicación a 10 a 19 en algunos participantes y en sentido inverso en otros, con la dirección contrabalanceada. La latencia media hasta la publicación fue cerca de un 10,9% mayor cuando la tasa era baja (0 a 9) que cuando era alta (10 a 19). Los autores advierten que hay muchos motivos para publicar además de la recompensa, que el estudio trata de cuándo se publica y no de lo que se publica, y que los datos no dicen si el uso intenso es desadaptativo.

Ceylan et al. (2023) estudiaron el acto de compartir. En cuatro estudios (N = 2.476), personas con cuenta en Facebook veían titulares (en los estudios 1 y 2, en el formato de Facebook) y decidían, mediante un botón simulado de compartir, si los compartían: eran 16 titulares, ocho verdaderos y ocho falsos en los estudios 1, 2 y 4. La fuerza del hábito de compartir noticias se midió con cuatro ítems de autoinforme, en una escala de 1 (nunca) a 7 (siempre), adaptados de un índice de hábito, como "compartir noticias en Facebook es algo que hago automáticamente"; la media de los ítems dio la puntuación de hábito de cada persona. En el estudio 1 (N = 200), en promedio se compartió el 32% de los titulares verdaderos y el 5% de los falsos. Quienes tenían el hábito más fuerte (una desviación estándar por encima de la media) compartieron pro-

porciones semejantes de titulares verdaderos (43%) y falsos (38%), y quienes lo tenían más débil (una desviación estándar por debajo), el 15% y el 6%; el 15% más habitual fue responsable del 37% de los falsos compartidos. En el estudio 2 (N = 839), los participantes fueron sorteados para elegir primero si compartían los 16 titulares y después juzgar la exactitud de cada uno, o para juzgar la exactitud de todos antes de decidir si los compartían. Juzgar la exactitud antes redujo el compartir de falsos, pero no aumentó la discriminación de los más habituales más que la de los demás.

En el estudio 4 (601 participantes, de los cuales 120 fueron excluidos por no superar la comprobación de que no había recompensa en la prueba, como estaba previsto en el prerregistro), cada persona fue sorteada para una de tres condiciones e hizo 80 ensayos de entrenamiento: en cada uno, pulsaba una tecla para compartir o no un titular y veía "ganó +5 puntos" o "ganó 0 puntos". En el entrenamiento de exactitud, ganaba 5 puntos al compartir un titular verdadero o al no compartir uno falso; en el entrenamiento de desinformación, a la inversa; en el control, no había retorno. A continuación decidía sobre 16 titulares nuevos, ocho verdaderos y ocho falsos, como lo haría en Facebook y sin recompensa. En la prueba, sin retorno, el patrón del entrenamiento persistió: después del entrenamiento de exactitud, se compartió el 66% de los verdaderos y el 24% de los falsos, frente al 40% y el 24% en el control y el 54% y el 33% en el entrenamiento de desinformación.

El hábito, en el uso de los autores, es la respuesta recompensada que se repite en un contexto, se asocia a señales recurrentes y pasa a ocurrir con poca sensibilidad a los resultados (Ceylan et al., 2023). En términos operantes, es una respuesta mantenida por reforzamiento y bajo control de estímulos recurrentes del contexto, aquí el formato de la publicación y el botón de compartir, que persiste cuando el reforzador sale de escena. Lo que el estudio 4 añade es que la veracidad puede disponerse como contingencia: con puntos para la exactitud, la diferencia entre verdaderos y falsos compartidos aumentó y persistió sin retorno, y el entrenamiento de desinformación elevó la frecuencia general de compartir sin derribar la discriminación de base. Los autores concluyen que el hábito surgió como la influencia principal en el compartir, por encima del sesgo político y de la reflexión crítica.

La monetización y los votos también son consecuencias para el productor, pero ningún estudio verificado mide ingresos o votos como resultado de publicar; quedan como hipótesis en la sección 9.7.

9.3 Lo que hace la audiencia y lo que la mantiene

La audiencia también se comporta, y es en ella donde nace el reforzamiento del productor. Lo que muestran los estudios verificados es, sobre todo, lo que hace ante cada tipo de contenido. Rathje et al. (2021) analizaron 2.730.215 publicaciones de Facebook y de Twitter. En el estudio 1, eran cuentas de medios de noticias clasificados como liberales (por ejemplo, New York Times y MSNBC) o conservadores (por ejemplo, Fox News y Breitbart), con 599.999 publicaciones de Facebook y 227.229 tuits; en el estudio 2, que replicó el primero en otro contexto, eran cuentas de miembros del Congreso de los Estados Unidos, demócratas y republicanos, con 1.078.562 tuits (hasta 3.200 de cada legislador) y 825.424 publicaciones de Facebook. Los datos se recogieron en 2020.

Para saber si una publicación hablaba del propio campo político (el endogrupo) o del opuesto (el exogrupo), los autores contaron, en cada texto, las palabras de tres listas: los 100 políticos demócratas y republicanos más famosos según YouGov, todos los legisladores de los dos partidos y unos diez términos de identidad política (como “liberal” y “conservative”). Un tuit con dos palabras de la lista republicana, por ejemplo, recibía una puntuación de 2 en esa categoría. En una cuenta conservadora, las palabras conservadoras contaban como lenguaje del endogrupo y las liberales, como lenguaje del exogrupo; en las cuentas liberales valía lo inverso. Con el mismo método contaron palabras de afecto negativo, de afecto positivo y de lenguaje moral-emocional, con diccionarios ya validados en otros estudios, para comparar el peso del lenguaje sobre el exogrupo con el de esas emociones. En cada conjunto de datos, una regresión relacionó el número de veces compartido (en logaritmo) con esos recuentos, descontados el efecto de tener enlace, foto o video, el número de seguidores de la cuenta y el hecho de que el tuit fuera un retuit; las mismas regresiones se aplicaron a cada reacción de Facebook. El resultado de cada regresión es un coeficiente por palabra, que los autores expresan como aumento porcentual.

Cada término referido al exogrupo se asoció con un aumento del 35 al 57% en la difusión en las cuentas de medios de noticias y del 65 al 180% en las del Congreso, según el conjunto. La media de los ocho conjuntos, en un metanálisis de efectos aleatorios hecho por los autores, fue de cerca del 67% por palabra (41% en los medios y 99% en el Congreso), un efecto cerca de 4,8 veces el del lenguaje de afecto negativo y 6,7 veces el del lenguaje moral-emocional. Los autores hablan de aumento de la probabilidad de que la publicación se comparta; el modelo estima, más exactamente, la razón entre los números esperados de veces compartido. Las publicaciones que se referían al exogrupo

se compartieron cerca de dos veces más que las que se referían al endogrupo. Facebook ofrece seis reacciones ("like", "love", "haha", "sad", "wow" y "angry"): la "angry" fue la más popular en siete de los ocho conjuntos, según el recuento de los autores, y el lenguaje sobre el exogrupo predijo reacciones "angry", y el lenguaje sobre el endogrupo, reacciones "love". El efecto del exogrupo no fue moderado por la orientación política ni por la plataforma, y fue más fuerte entre líderes políticos que entre medios de noticias.

Los autores leen el resultado en términos de identidad social: los mensajes que satisfacen motivos de identidad de grupo pueden recibir más interacción, y las publicaciones sobre el exogrupo pueden tener éxito por apelar a la ira, la indignación y la burla (Rathje et al., 2021). Es una lectura de otro marco de referencia. En términos operantes, lo que el estudio establece es que ciertas propiedades del contenido, aquí la referencia al exogrupo, se asocian con más respuestas de compartir y de reaccionar; por qué esas respuestas se refuerzan para quien las emite es lo que no mide.

Guerin (1994) ofrece una hipótesis vecina para esta parte del ciclo. Intercambiar actitudes puede funcionar como evento social ritual, en el que el reforzamiento está en el propio intercambio y no en lo que se dice, y la audiencia es estímulo discriminativo, de modo que lo que se declara tiende a atender a las contingencias de la comunidad que ocasionó la declaración. Para la audiencia de una red, esto sugiere que dar "me gusta", compartir o comentar puede ser, antes que una evaluación del contenido, parte del intercambio que mantiene a la comunidad. Es hipótesis: los estudios citados no miden la consecuencia que recibe la propia audiencia.

DATO	La tasa de "me gusta" se asocia con un menor intervalo entre publicaciones (Lindström et al., 2021); el patrón de compartir, instalado por retorno, persiste cuando el retorno se retira (Ceylan et al., 2023); las publicaciones sobre el exogrupo político se comparten cerca de dos veces más que las publicaciones sobre el endogrupo (Rathje et al., 2021).
INTERPRETACIÓN	Publicar y compartir son operantes sensibles al retorno social y al contexto de la plataforma, y el contenido al que la audiencia más responde es, en parte, lo que el productor pasa a producir.
HIPÓTESIS	El compartir y el reaccionar de la audiencia se mantienen por el retorno que ella recibe de la propia red. Ninguno de los estudios lo mide.

9.4 El ciclo: cada lado moldea al otro

El ciclo es el encaje de dos contingencias de tres términos, la del productor (sección 9.2) y la de la audiencia (sección 9.3). La del productor tiene como ocasión la audiencia, la norma de la red y la interfaz de la plataforma; como respuesta, publicar un contenido; como consecuencia, "me gusta", veces compartido y alcance. La de la audiencia tiene como ocasión el contenido que aparece en el feed; como respuesta, dar "me gusta", compartir, comentar o ignorar; y, como consecuencia, lo que los estudios no miden y la sección 9.3 dejó como hipótesis. Lo que las hace un ciclo es que cada una contiene, como término, la respuesta de la otra: la publicación es la ocasión de la audiencia, y la reacción de la audiencia es la consecuencia del productor. Es lo que Glenn (2004) llama contingencias conductuales entrelazadas.

CONCEPTO EN USO

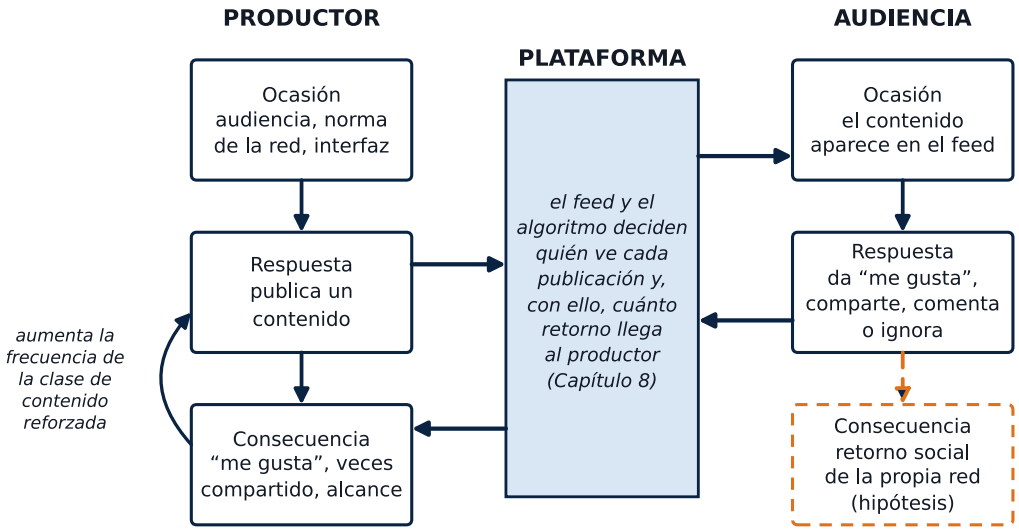
Contingencias conductuales entrelazadas (Glenn, 2004), retomado del Capítulo 5

El Capítulo 5 (sección 5.1) presentó el concepto: las contingencias están entrelazadas cuando la conducta de dos o más personas funciona como evento ambiental para la conducta de las otras (Glenn, 2004), y el resultado no es el efecto acumulativo de cada persona comportándose sola, sino el de su conducta interrelacionada. En el Capítulo 5 los participantes eran hablante, audiencia y grupo destinatario; aquí son el productor y la audiencia, y entre ellos está la plataforma.

La figura no es un hallazgo, es un modelo. Dos de sus flechas tienen datos en los estudios del capítulo: la que va del retorno a la producción (Lindström et al., 2021; Ceylan et al., 2023; Brady et al., 2021, sección 9.5) y la que une el tipo de contenido con la reacción de la audiencia (Rathje et al., 2021). La que cierra el ciclo, en la que lo que la audiencia ve cambia lo que pasa a hacer, tiene apoyo indirecto: Brady et al. (2021) muestran, en un entorno simulado, que lo que aparece en el feed funciona como norma y que la norma altera la expresión (sección 9.5). Ningún estudio mide, en la misma población y a lo largo del tiempo, las dos mitades. La franja del medio es la plataforma, que decide quién ve qué y, por lo tanto, cuánto retorno llega a quien publica; las fuentes verificadas dicen poco sobre ella, y el Capítulo 8 trata de ella. Aquí es el término medio.

FIGURA 9.1

El ciclo productor–audiencia: dos contingencias de tres términos que se entrelazan



La publicación es la ocasión de la audiencia; la reacción de la audiencia es la consecuencia del productor. Línea continua: relación con datos en los estudios del capítulo. Discontinua: hipótesis.

Nota. Esquema original del autor, sin datos. Línea continua: relaciones con datos en los estudios del capítulo; discontinua: hipótesis. La plataforma decide cuánto retorno llega al productor (Capítulo 8).

9.5 Indignación moral aprendida por reforzamiento: la evidencia

El estudio que más se acerca al ciclo de la sección 9.4, entre los leídos, es el de Brady et al. (2021), sobre la expresión de indignación moral. Los autores parten de dos hipótesis: que las personas cambian lo que expresan por aprendizaje por reforzamiento, alterando conductas expresivas en respuesta al retorno social, y por aprendizaje de normas, ajustando la expresión a lo que infieren que es normativo entre sus pares, por observación. El diseño tiene dos partes.

En la primera, dos estudios observacionales prerregistrados reunieron los historiales completos de tuits de usuarios “políticamente comprometidos”, identificados por haber tuiteado al menos una vez sobre las audiencias de confirmación de un juez de la Corte Suprema, en octubre de 2018 (estudio 1, 3.669 usuarios), o sobre un incidente en el que una aerolínea maltrató a un pasajero (estudio 2, la misma cantidad); los autores cuentan 7.331 usuarios en total, con 12,7 millones de tuits. Para decir si un tuit expresaba indignación moral, la definieron por tres componentes (sentimientos ante la violación percibida de principios morales personales; emociones como ira, asco y desprecio; y reacciones como culpar, responsabilizar o querer castigar) y entrenaron un clasifi-

cador con 26.000 tuits anotados con base en esa definición. El clasificador tuvo una exactitud del 75% (proporción de tuits clasificados correctamente) y una *F*1 de 0,71 (puntuación de 0 a 1 que combina cuántos de los tuits de indignación encontró y cuántos de los que marcó como indignación lo eran realmente).

Después, para cada usuario, relacionaron la cantidad de indignación de un día con el retorno ("me gusta" y veces compartido) que había recibido la indignación del día anterior, descontados la frecuencia de tuitear, el número de seguidores, la presencia de enlaces y de contenido multimedia, el historial de la semana anterior y el retorno que no era específico de la indignación (para separar el efecto del retorno a la indignación del simple hecho de tuitear más cuando se recibe más retorno). La expresión diaria de indignación se asoció positivamente con el retorno recibido por la indignación del día anterior (coeficientes de regresión *b* de 0,03 y 0,02 en los estudios 1 y 2). En el ejemplo de los autores, quien tenía en promedio cinco "me gusta" o veces compartido por tuit y recibió diez al expresar indignación (un aumento del 100%) tendería a aumentar la indignación del día siguiente en un 2 a 3%, un efecto que los autores llaman pequeño y capaz de escalar. Cuanto más largo el historial del usuario, menor el efecto del retorno reciente, y los usuarios insertos en redes ideológicamente más extremas expresaron más indignación y respondieron menos al retorno.

En la segunda parte, dos experimentos prerregistrados (120 participantes cada uno, reclutados en Prolific, todos liberales) simularon un Twitter. En el estudio 3, los participantes eran sorteados para una de dos condiciones de norma y empezaban desplazándose por un feed de 12 tuits de una "nueva red", cuyas preferencias debían aprender. En la condición de norma de indignación, el 75% de los tuits expresaba indignación y exhibía más "me gusta" y veces compartido que los neutros; en la de norma neutra, todos eran neutros. A continuación hacían 30 ensayos en los que elegían, entre dos tuits políticos, uno de indignación y uno neutro, para "publicar" en la red, y recibían "me gusta" que, según se les decía, venían de participantes anteriores; el objetivo era maximizar esos "me gusta". Elegir indignación rendía más "me gusta", en promedio, en las dos condiciones. El estudio 4 usó el mismo paradigma con una diferencia: el retorno mayor iba para el tuit congruente con la norma, es decir, para la indignación en la condición de norma de indignación y para el tuit neutro en la de norma neutra.

Los autores midieron dos procesos. El aprendizaje de normas es la tendencia a elegir, en el primer ensayo, el tipo de tuit que el feed mostraba como normal. El aprendizaje por reforzamiento es la tendencia a elegir cada vez más, a lo largo de los ensayos, el tipo de tuit que rendía más "me gusta". Los resultados

vienen como razones de probabilidades (OR, de *odds ratio*): una OR de 1 significa probabilidades iguales, y por encima de 1, ventaja para la elección en cuestión. En el primer ensayo, quienes vieron la norma de indignación eligieron el tuit de indignación con una probabilidad mucho mayor (OR = 4,94 en el estudio 3 y 5,38 en el estudio 4), y quienes vieron la norma neutra eligieron el neutro (OR = 1,73 y 1,54). A lo largo de los ensayos, la elección del tuit reforzado aumentó (en el estudio 3, OR = 1,10 en el conjunto de las dos condiciones; en el estudio 4, 1,03 para indignación y 1,06 para tuits neutros). Y el efecto del reforzamiento fue menor entre quienes vieron la norma de indignación (OR de 0,95 en el estudio 3 y de 0,97 en el estudio 4 para la diferencia entre las condiciones; por debajo de 1, el efecto es menor en esa condición), lo que los autores leen como menor dependencia del retorno para quienes ya veían la indignación como norma.

Los términos de los autores vienen del aprendizaje por reforzamiento computacional (retorno, error de predicción). Lo que importa aquí es la relación funcional: la expresión de indignación varió con las consecuencias sociales que la antecedieron, en datos naturales y en un entorno en el que el retorno se dispuso. Lo que el reforzamiento por sí solo no explica, dicen los autores, es que el feed muestra lo que es normativo en la red y quien lo observa depende menos del retorno. En lenguaje analítico-conductual, esto se acerca al control por modelo y por regla (lectura aquí propuesta), y no es lo mismo que el reforzamiento directo. El vínculo con los algoritmos es interpretación de los autores: al determinar cuántos usuarios ven una publicación, los algoritmos afectan el retorno y pueden influir en conductas morales al explotar tendencias de aprendizaje por reforzamiento; y las decisiones de diseño orientadas al lucro por interacción pueden afectarlo indirectamente, porque el contenido que provoca indignación atrae una interacción alta. Ninguno de los cuatro estudios mide decisiones de diseño.

Los propios autores registran los límites: los estudios observacionales no permiten inferencia causal, los usuarios fueron identificados por haber tuiteado en episodios de indignación pública, y las conclusiones se refieren a la expresión en texto, y no a la emoción (sección 9.8).

DATO	En 12,7 millones de tuits de 7.331 usuarios, la indignación del día siguiente aumentó del 2 al 3% por cada 100% de aumento del retorno a la indignación del día anterior; en dos experimentos de 120 participantes, la elección de tuits de indignación aumentó con el retorno y siguió la norma exhibida (Brady et al., 2021). En más de un millón de publicaciones, el intervalo entre publicaciones disminuyó con la tasa de “me gusta” (Lindström et al., 2021).

INTERPRETACIÓN	La expresión de indignación, como la frecuencia de publicar, es sensible al retorno social; el contexto de la red, observado en el feed, modula esa sensibilidad.
HIPÓTESIS	En las plataformas en las que el algoritmo amplía lo que provoca interacción, el retorno por la indignación crece y la expresión crece con él. Es hipótesis de los autores, retomada aquí; ningún estudio manipuló el algoritmo y midió la indignación.

9.6 El político como productor de contenido

Un político que publica en una red es un productor con una audiencia particular: seguidores con su propia historia de reforzamiento sobre el tema, electores potenciales, prensa. Rathje et al. (2021) separan el análisis por tipo de cuenta y por campo político, lo que permite leer el ciclo para este productor. En 1.078.562 tuits y 825.424 publicaciones en Facebook de congresistas demócratas y republicanos, las que se referían al exogrupo se compartieron más en los dos campos. El aumento estimado por término sobre el exogrupo, en el metanálisis de los autores, fue del 99% en los conjuntos de congresistas, frente al 41% en los de medios de noticias, y el efecto no fue moderado por la orientación política. El estudio cubre los años que precedieron a la elección de 2020, es correlacional, y los autores dicen no saber cómo contribuyen las elecciones algorítmicas.

El hallazgo es sobre cuentas y contenido: dice qué se compartió más, y no por qué cada político publicó lo que publicó. El estudio aplica el mismo criterio a los dos campos, que es lo que pide el método seguido aquí. Aplicarlo a políticos brasileños es la pregunta que se quiere hacer; ninguna de las fuentes verificadas la responde.

DATO	Las publicaciones de congresistas de los Estados Unidos que se referían al exogrupo político se compartieron más que las que se referían al endogrupo, en Facebook y en Twitter, en los dos campos, con un efecto mayor que el observado en medios de noticias (Rathje et al., 2021).
INTERPRETACIÓN	El retorno que la audiencia da a un productor político favorece, en los datos, el contenido sobre el exogrupo, lo que describe contingencias del productor y no juzga a personas ni a partidos.
HIPÓTESIS	Lo mismo vale para los políticos brasileños y para otras plataformas. Ninguna fuente verificada lo pone a prueba.

9.7 La metacontingencia del ecosistema: productor, audiencia, plataforma y anunciante

Hasta aquí el capítulo se quedó en el nivel de la conducta de individuos en contingencias entrelazadas. La pregunta que queda es si el conjunto, formado por productores, audiencias, plataformas y quienes pagan por atención, se mantiene por sí mismo, como una práctica que se repite porque sus efectos la seleccionan. El Análisis de la Conducta tiene un concepto para eso, la metacontingencia. El capítulo lo usa como hipótesis y con los cuidados que pide la propia literatura.

CONCEPTO EN USO

Metacontingencia (Glenn, 1988, 2004)

La metacontingencia es la unidad de análisis que abarca una práctica cultural, en todas sus variaciones, y el resultado agregado de las variaciones vigentes (Glenn, 1988). Los resultados culturales no seleccionan la conducta de individuos; seleccionan las contingencias conductuales entrelazadas que componen la práctica. En Glenn (2004), los linajes que evolucionan son esas contingencias recurrentes, que funcionan como una unidad integrada y tienen un resultado que afecta la probabilidad de su recurrencia futura. La fórmula "contingencias conductuales entrelazadas, su producto agregado y su sistema receptor" es de Glenn y Malott (2004, citados por Houmanfar & Rodrigues, 2006). La macrocontingencia es otra cosa: la relación entre una práctica y la suma agregada de las consecuencias del macrocomportamiento que la constituye, efectos acumulativos que no pueden controlar, como consecuencias operantes, la conducta de individuos (Glenn, 2004) y que no implican selección en el nivel cultural.

Aplicado al ecosistema, el esquema exige tres cosas: contingencias entrelazadas recurrentes, un producto agregado y un sistema receptor que, al recibir el producto, mantiene la recurrencia de lo que lo generó. La Figura 9.2 propone cómo completarlo. Las contingencias entrelazadas son las de la sección 9.4, con la plataforma como tercer término: el productor publica y recibe reacciones, la audiencia ve y reacciona, la plataforma ordena el feed y retiene atención. El producto agregado sería un patrón recurrente de contenido y de interacción, por ejemplo publicaciones sobre el exogrupo muy compartidas. El sistema receptor sería quien paga por la atención retenida: los anunciantes y, por medio de ellos, los ingresos de la plataforma.

FIGURA 9.2

Hipótesis: la metacontingencia del ecosistema productor–audiencia–plataforma



Nota. Esquema original del autor, sin datos: es una hipótesis, no un resultado. Los tres términos siguen la fórmula de Glenn y Malott (2004, citados por Houmanfar & Rodrigues, 2006). Houmanfar y Rodrigues proponen otra configuración: ambiente cultural, producto agregado, sistema receptor.

Las fuentes verificadas sostienen las contingencias entrelazadas en parte (secciones 9.2 a 9.5) y dejan la segunda mitad de la hipótesis como supuesto de sus autores: Brady et al. (2021) afirman que el contenido que provoca indignación atrae una interacción alta, lo que puede afectar la conducta moral cuando el diseño busca el lucro por interacción; Rathje et al. (2021), que el modelo de negocio de las redes se basa en generar interacción, lo que lleva a los ingresos por publicidad; Ceylan et al. (2023), que los usuarios habituales son parte integral de los modelos de lucro basados en anuncios. Son premisas de interpretación, y no resultados: ninguno de esos estudios midió ingresos ni decisiones de diseño.

Dos cuidados vienen de la propia literatura. El primero es la distinción entre metacontingencia y macrocontingencia. Si lo que se quiere explicar es un efecto acumulativo, como la polarización (ejemplo aquí elegido), lo que hay es una macrocontingencia: el efecto agregado de muchos operantes independientes no puede funcionar como consecuencia que controla la conducta de cada individuo (Glenn, 2004). Lo que mantiene al usuario publicando sigue siendo el retorno que él mismo recibe. La ingeniería cultural, escribe Glenn (2004), exige identificar las macrocontingencias que producen efectos indeseables y alterar las contingencias operantes o las metacontingencias relevantes; la Parte IV vuelve sobre esto. El segundo cuidado es que el concepto está en discusión. Houmanfar y Rodrigues (2006) proponen la configuración ambien-

te cultural, producto agregado y sistema receptor, con el producto agregado como lo que se selecciona, porque consideran que la forma actual mezcla niveles de análisis, y dicen que las ideas son preliminares. Para el ecosistema la diferencia importa: si lo que el sistema receptor selecciona es el producto agregado, y no las contingencias entrelazadas, la pregunta empírica es qué productos se repiten (qué tipo de contenido y de interacción), y no qué arreglos de productor, audiencia y plataforma.

DATO	Los estudios verificados muestran sensibilidad al retorno social al publicar, compartir y expresar indignación (secciones 9.2 a 9.5) y que el contenido sobre el exogrupo se comparte más (9.3 y 9.6). Sus autores afirman, sin medirlo, que los modelos de negocio de las plataformas dependen de la interacción.
INTERPRETACIÓN	Las contingencias del productor y de la audiencia están entrelazadas en el sentido de Glenn (2004), con la plataforma como mediadora del retorno.
HIPÓTESIS	El patrón entrelazado se repite porque su producto agregado lo reciben anunciantes que pagan por atención. Eso exigiría mostrar que a la variación del producto agregado le sigue una variación en los ingresos y que estos mantienen la práctica. Ninguna fuente verificada lo hace.

9.8 Interpretaciones rivales y límites

El ciclo tiene lecturas rivales, y lo que separa una de otra depende del tipo de dato. Los datos de plataforma son de dos tipos, con límites diferentes. Los observacionales (Lindström et al., 2021; Brady et al., 2021, estudios 1 y 2; Rathje et al., 2021) describen asociaciones en poblaciones autoseleccionadas; los experimentales (Lindström et al., 2021, estudio 3; Brady et al., 2021, estudios 3 y 4; Ceylan et al., 2023) manipulan el retorno en entornos simulados, con participantes de plataformas de investigación, y miden elecciones, no publicaciones en redes reales.

Cuatro límites valen para todo el capítulo. Primero, ninguna fuente es brasileña, y las de plataforma son de los Estados Unidos o de muestras de investigación en línea; Brady et al. (2021) dicen no saber si los hallazgos se generalizan a otras poblaciones, a otras plataformas o fuera del contexto político de los Estados Unidos, y que los usuarios de Twitter no son representativos. Segundo, los efectos medidos son pequeños: del 2 al 3% de aumento en la indignación del día siguiente por cada 100% de aumento del retorno (Brady et al., 2021) y una diferencia de cerca del 10,9% en la latencia (Lindström et al.,

2021). Los efectos pequeños pueden escalar, como señalan los autores, pero eso es un argumento, y no un resultado. Tercero, los experimentos de 1956 a 1969 usan estudiantes y, en Scott (1957), un intervalo de cinco minutos. Cuarto, el ciclo como un todo es un modelo: cada flecha tiene apoyo parcial, y el cierre del ciclo a lo largo del tiempo, con productores y audiencias reales, no se ha medido.

Interpretaciones rivales para el mismo conjunto de datos

Exposición determinada por el algoritmo. Lo que recibe una publicación depende de cuántos usuarios la ven, y eso lo decide la plataforma (Brady et al., 2021). Una asociación entre contenido y reacciones puede reflejar lo que el algoritmo mostró, y no lo que la audiencia elegiría; Rathje et al. (2021) dicen no saber cómo contribuyen las elecciones algorítmicas.

Selección del productor y de la red. Los usuarios que expresan más indignación pueden seguir más cuentas extremas, en lugar de aprender a expresarla en la red en la que están; los autores admiten que la composición de la red puede ser endógena (Brady et al., 2021, estudios 1 y 2).

Otros motivos para publicar. Hay muchos motivos además de la recompensa, como la autoexpresión y el desarrollo de relaciones (Lindström et al., 2021).

Norma e identidad en lugar de reforzamiento. Lo que se observa puede ser conformidad con las normas de la red (Brady et al., 2021) o expresión de identidad de grupo (Rathje et al., 2021), sin que el retorno individual sea lo que mantiene la respuesta.

Expresión sin emoción. Las conclusiones se refieren a la expresión en texto, y no a la emoción, que no se midió; los autores admiten que el diseño de las plataformas puede desacoplar una de otra (Brady et al., 2021).

Del teléfono a la red. Los experimentos de 1956 a 1969 usan un oyente conocido, una reacción verbal y pocas interacciones; las redes tienen audiencias múltiples, retorno contado y visible y ciclos largos. El paso es por analogía, y aquí se asume como tal.

Para quien quiera usar el capítulo en el análisis de declaraciones (Capítulos 3, 5 y 7), la consecuencia es práctica. La reacción de quien oye es una variable de control plausible y medible de lo que dice el hablante, pero afirmar que la declaración de un político específico fue mantenida por determinada audiencia exige el registro de esa audiencia y de sus reacciones, que una verificación, por sí sola, no aporta.

Síntesis

Lo que se dice es sensible a lo que ocurre después de decirlo, y esto se mostró con opiniones declaradas antes de que existieran las redes sociales: con "Good" después de respuestas, con un voto de victoria después de un debate. El efecto es pequeño, aparece fuera del laboratorio, depende del reforzador y de la tarea, y la percepción de la contingencia está en disputa (sección 9.1). En las plataformas, la frecuencia de publicar sigue la tasa de "me gusta", el compartir se vuelve poco sensible a la veracidad cuando se instala por repetición y la expresión de indignación sigue el retorno y la norma exhibida en el feed, con efectos pequeños en los datos observacionales (secciones 9.2 y 9.5). Del lado de la audiencia, se sabe lo que hace ante cada contenido; la consecuencia que la mantiene es hipótesis (9.3). El ciclo une las dos mitades como contingencias entrelazadas, tiene apoyo parcial en cada flecha y ningún estudio lo mide por entero (9.4).

Para el político como productor, el único dato verificado es el de los congresistas de los Estados Unidos: el contenido sobre el exogrupo se comparte más, en los dos campos, y más entre congresistas que entre medios (9.6). El paso a Brasil es hipótesis. En el nivel cultural, el ecosistema puede leerse como metacontingencia, con la plataforma y los anunciantes como términos, pero eso exige un resultado que ninguna fuente verificada mide, y parte de lo que se le atribuye, como la polarización, es un efecto acumulativo, de macrocontingencia (9.7).

Los Capítulos 3, 5 y 7 toman de este capítulo la idea de que la reacción de la audiencia es una variable de control de la declaración; el Capítulo 8, la franja de la plataforma; el Capítulo 12, la distinción entre actuar sobre las contingencias operantes del productor y de la audiencia y actuar sobre las metacontingencias.

Referencias

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), Artículo 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), Artículo 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), Artículo 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), Artículo 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

instituto WALDEN4 | ACADEMIA WALDEN4

TV WALDEN4

Ciencia en video, al aire todo el día

Clases, conferencias y conversaciones sobre psicología, autismo y Análisis de la Conducta, en canales temáticos que se transmiten sin parar, sincronizados para todo el mundo.

667 videos en 10 canales

Acceda, gratis y sin registro

walden4.com.br/tv

CAPÍTULO

10

La ventana de Overton: lo que se puede decir sin ser castigado

Un experimento de Bursztyn et al. (2020), realizado en 2018, sirve de entrada a la ventana de Overton. Los 1.600 residentes del área metropolitana de Pittsburgh que participaron recibieron la misma oferta: una bonificación en dinero si autorizaban una donación, en su nombre, a una organización fuertemente antiinmigración. Al azar, unos supieron que Trump había ganado en el área metropolitana en la elección de 2016; otros, que Clinton se había llevado el condado en el que está Pittsburgh. A la mitad de cada grupo se le hizo creer que la decisión sería anónima; a la otra mitad, que sería visible para otros residentes de la región. En la condición anónima, aceptaron el 31% (ganó Trump) y el 30% (ganó Clinton). En la condición visible, el 33% y el 20% (Bursztyn et al., 2020).

La información que variaba era una sola y no decía nada sobre inmigración: decía quién había ganado la elección en el vecindario. Lo que varió con ella fue lo que las personas aceptaron hacer cuando otras podían verlas. Es ese desplazamiento de lo que se puede hacer o decir en público sin costo lo que la conversación política llama desplazamiento de la ventana de Overton.

El capítulo pregunta qué significa el concepto en términos de contingencias y qué permiten decir los datos verificados sobre quién desplaza la ventana y por qué. La sección 10.1 presenta el concepto, el origen y la traducción aquí propuesta; la 10.2 lo sitúa en la relación entre hablante y audiencia; la 10.3 vuelve al experimento de Pittsburgh; la 10.4 trata de dos elecciones, la brasileña de 2018 y la estadounidense de 2016; la 10.5, de políticos y redes sociales; la 10.6, de lo que mantiene el borde en su lugar; la 10.7 propone preguntas para situar un dicho; y la 10.8 reúne las interpretaciones rivales.

Tres cautelas al leer el capítulo. La primera es que el concepto nació en un centro de investigación que lo usa para promover sus propias políticas y es, en su origen, una metáfora sobre políticas públicas, y no una definición de conducta verbal. La segunda es que los datos miden cosas diferentes (percepción, acción pública, texto clasificado), y ningún estudio verificado vincula el discurso de un político con el de sus electores mediante un diseño que identifique la dirección de la influencia. La tercera es que los estudios de elección tratan de las victorias de Trump, en 2016, y de Bolsonaro, en 2018; ninguno trata de un caso especular, la victoria de un candidato del campo opuesto. El criterio del capítulo vale para cualquier campo, pero lo que se ha mostrado no permite afirmar simetría. En todas las secciones, dato, interpretación e hipótesis quedan separados.

10.1 El concepto, el origen y la traducción propuesta

La formulación de origen es la del Mackinac Center for Public Policy, que mantiene una página sobre el concepto. Según ella, la ventana de Overton es un modelo para entender cómo las ideas de la sociedad cambian con el tiempo e influyen en la política. El núcleo es que los políticos están limitados en lo que pueden apoyar: en general solo adoptan políticas ampliamente aceptadas como legítimas, y esas están dentro de la ventana; las demás existen, pero quien las defiende se arriesga a perder apoyo popular, y están fuera de ella (Mackinac Center, 2019).

La ventana puede desplazarse y expandirse. A veces los políticos la mueven al respaldar una política que está fuera de ella, pero eso es raro; con más frecuencia la ventana se mueve por la lenta evolución de los valores y de las normas de la sociedad, que no se controla fácilmente desde arriba. La conclusión de la página es que los políticos son más seguidores que líderes, y que las instituciones que establecen y refuerzan normas (familias, lugares de trabajo, amigos, medios, iglesias, asociaciones, escuelas) importan más para moldear la política de lo que se suele reconocer. El ejemplo es la Ley Seca: política federal hace algunas generaciones, hoy ningún político defiende restablecerla, y salió de la ventana. Interesa aquí un pasaje de las preguntas frecuentes: la ventana refleja lo que la sociedad cree, y eso puede estar tan influido por hechos como por información imprecisa o engañosa; la página responde que la ventana puede ser desplazada por mentiras, distorsiones o malentendidos, y añade que es obviamente incorrecto diseminar intencionalmente información engañosa, y da como ejemplos de errores que desplazaron la ventana la subestimación de los costos de Medicare y la creencia en las armas de destrucción masiva de Irak (Mackinac Center, 2019).

El lector debe leer esto sabiendo de dónde viene. El concepto fue desarrollado a mediados de la década de 1990 por Joseph P. Overton, que era vicepresidente sénior del centro cuando murió, en 2003, y el centro dice usar la ventana de Overton para ayudar a promover sus propias políticas. La página admite, por último, que el modelo no describe todo sobre cómo funciona la política: describe una cosa, que los políticos solo defienden políticas que juzgan que no perjudican sus posibilidades electorales (Mackinac Center, 2019). La descripción no queda invalidada por eso, pero es de una parte interesada y no de un estudio, y el capítulo la usa como punto de partida, y no como evidencia.

Como formulación para una ciencia de la conducta, al concepto le falta algo y requiere un cambio. La falta es operacional: la página no dice cómo medir "ampliamente aceptada" ni "apoyo popular", de modo que, tal como está formulado, lo que está dentro o fuera de la ventana solo se conoce después, por lo que los políticos hicieron. El cambio es de unidad y de criterio. La unidad deja de ser la política pública que un político puede apoyar y pasa a ser el enunciado que un hablante puede emitir. El criterio deja de ser el "apoyo popular" y pasa a ser la consecuencia que el enunciado recibe de las audiencias que importan al hablante, para la cual el análisis de la conducta verbal ya tiene un concepto: la audiencia que castiga. El recuadro reúne lo que dice Skinner (1957).

CONCEPTO EN USO

Audiencia negativa y audiencia no punitiva (Skinner, 1957)

Skinner (1957) distingue al oyente que simplemente no refuerza la conducta verbal de aquel que la castiga: una audiencia en cuya presencia la conducta verbal es castigada puede llamarse "audiencia negativa", y reyes, altos funcionarios y ejecutivos poderosos pueden ser audiencias negativas efectivas. Hay audiencias negativas para asuntos (se aprende a no mencionar ciertos temas o eventos) y audiencias que controlan solo parte del repertorio del hablante, como la comunidad que castiga la lengua hablada en el hogar del niño. El efecto de una audiencia negativa solo se detecta en combinación con una audiencia positiva, porque aparece como reducción de la fuerza de la conducta apropiada para esta; cuando las dos están juntas, se instala un dilema, en el que la respuesta es emitida y castigada por una o retenida, con pérdida de prestigio ante la otra, y mientras las audiencias permanecen separadas el dilema no aparece. El orador sedicioso que ve acercarse a un policía habla menos a medida que la audiencia negativa se vuelve relativamente más importante, tal vez incluso por debajo del nivel abierto. Lo opuesto es la audiencia no punitiva: el psicoterapeuta necesita establecerse como tal, y, si el control de la audiencia cambia, el paciente puede emitir conducta antes castigada; el confidente es la persona comprensiva a quien se habla con menos miedo al castigo que a oyentes cualesquiera.

De ahí se toma la definición usada en el resto del capítulo. Se propone aquí, y no es de Skinner ni del Mackinac Center.

DEFINICIÓN PROPUESTA

Ventana de una audiencia.

Para un hablante y un asunto, el conjunto de enunciados que ese hablante emite en presencia de esa audiencia sin que ella los castigue. "Castigar" tiene el sentido funcional de Skinner: es la consecuencia que vuelve el enunciado menos probable, lo que incluye la pérdida de prestigio, y no solo la sanción formal.

De la definición se siguen dos consecuencias. La primera es que la ventana es propiedad de la relación entre hablante, audiencia y asunto, y no del enunciado: el mismo enunciado está dentro de la ventana de una audiencia y fuera de la de otra, y no hay "la" ventana de una sociedad, hay tantas como audiencias relevantes. La ventana de la que habla la conversación política, la del "debate público", es, en la lectura aquí propuesta, la de las audiencias que dominan una arena en un momento, y el análisis necesita decir cuáles. La segunda es que "dentro" y "fuera" no son categorías de lo que es verdadero, justo o correcto: dicen lo que suele o no suele ser castigado por quien escucha. Un enunciado falso puede estar dentro de la ventana de una audiencia, y uno verdadero, fuera de la de otra. La definición no toma partido sobre ninguno de los dos.

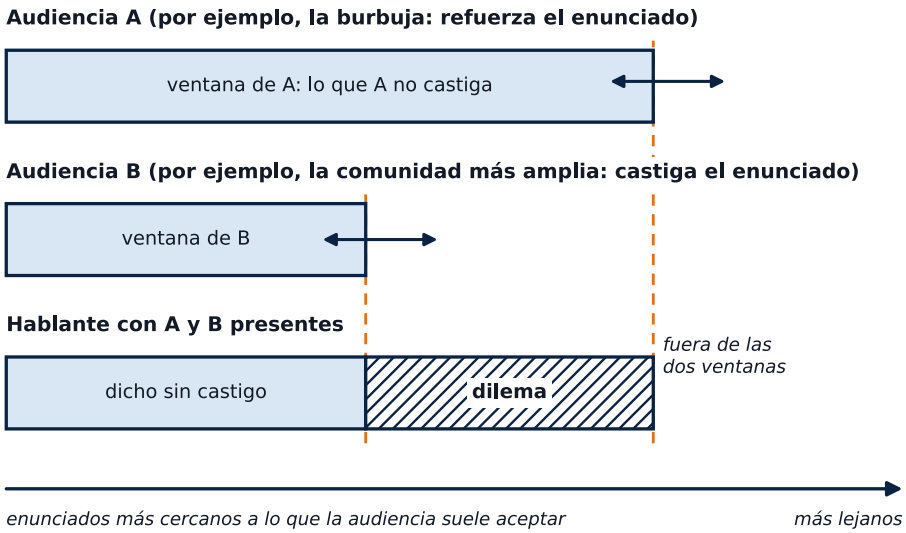
DATO	El Mackinac Center describe la ventana como el conjunto de políticas que los políticos pueden apoyar sin perder apoyo popular, dice que se mueve sobre todo por la lenta evolución de valores y normas y que los políticos son más seguidores que líderes; el propio centro dice usar el concepto para promover sus políticas (Mackinac Center, 2019). Skinner (1957) describe la audiencia negativa, el dilema de audiencias con ventanas diferentes y la audiencia no punitiva.
INTERPRETACIÓN	La ventana de una audiencia es el conjunto de enunciados que ella no castiga, para un hablante y un asunto; es relación, y no propiedad del enunciado.
HIPÓTESIS	Los políticos "siguen" en la medida en que lo que emiten está controlado por el castigo de las audiencias que les importan. Ponerla a prueba exige medir la sanción, y no solo lo que se dijo; la página del Mackinac Center no dice cómo.

10.2 De la ventana de la política a la ventana del enunciado: lo que desplaza el borde

La definición de ventana como propiedad de la relación entre hablante, audiencia y asunto (sección 10.1) tiene una implicación que la Figura 10.1 dibuja. Dos audiencias tienen ventanas de anchura diferente para el mismo asunto: la de A, por ejemplo una burbuja (Capítulo 6) en la que el enunciado es reforzado, es ancha, y la de B, por ejemplo una comunidad más amplia, es estrecha. Cuando solo A está presente, el hablante emite hasta el borde de A; cuando solo está B, hasta el de B. Cuando las dos están presentes, la franja entre los bordes es la zona de dilema de Skinner: el enunciado es reforzado por A y castigado por B, y el hablante lo emite y es castigado por B o lo retiene y pierde prestigio ante A. Lo que hace depende del peso relativo de cada audiencia, que la figura no representa.

FIGURA 10.1

La ventana es relativa a la audiencia: dos ventanas y la zona de dilema



Esquema conceptual original, sin datos. Flechas dobles: el borde se desplaza (sección 10.2). Rayado: reforzado por A y castigado por B (Skinner, 1957, pp. 230–231). Simplificación: las dos ventanas suponen el mismo orden de enunciados.

Nota. Esquema original, sin datos. Las flechas dobles indican que el borde se desplaza (véase el texto). La franja rayada es la zona de dilema (Skinner, 1957). Simplificación: las dos ventanas suponen el mismo orden de enunciados, del más al menos aceptado; en la realidad, dos audiencias pueden ordenar los enunciados de modos diferentes.

En términos de contingencias, el borde de una ventana se desplaza cuando cambia la probabilidad de castigo de los enunciados vecinos a él. Se reúnen aquí siete candidatas a explicar el cambio, cada una con lo que las fuentes verificadas aportan. La primera es quién forma parte de la audiencia: la entrada

o la salida de oyentes con otra historia cambia lo que es reforzado y lo que es castigado, y el Capítulo 6 trató las burbujas como audiencias que refuerzan lo que el grupo acepta. La segunda es el poder de sanción de la audiencia: Skinner (1957) habla de audiencias negativas "efectivas", y el contracontrol (Delprato, 2002), tratado en la sección 6.4, es el conjunto de respuestas que reduce la eficacia del control aversivo; sin poder de sanción, el borde retrocede. La tercera es la información sobre lo que los otros toleran, el objeto de las secciones 10.3 y 10.4. La cuarta es lo que el ambiente muestra que los otros hacen, esto es, las normas descriptivas, tratadas en la sección 10.6. La quinta es lo que recibe retorno en el borde: los enunciados limítrofes que son reforzados pasan a emitirse con más frecuencia, y la sección 10.5 trata de eso con datos de políticos. La sexta es la repetición: encontrar repetidamente un titular falso redujo lo poco ético que parecía publicarlo y compartirlo, en cuatro experimentos (Effron y Raj, 2020), resultado que el Capítulo 3 (sección 3.5) trata con más detenimiento. La séptima es el alcance: la plataforma decide cuántos oyen cada emisión (Capítulo 8), y una ventana ancha de una audiencia pequeña y una estrecha de una audiencia enorme tienen efectos muy diferentes sobre lo que circula.

DATO	Cada una de las variables anteriores tiene apoyo en al menos un estudio verificado, de naturalezas muy diferentes (experimentos, series de tweets, un experimento con repetición de titulares), y ninguna se manipuló junto con las otras.
INTERPRETACIÓN	El mismo cambio de discurso puede ser producido por más de una de ellas, y el término "desplazamiento de la ventana" describe el resultado, y no el mecanismo.
HIPÓTESIS	Las siete variables actúan en conjunto en la práctica (por ejemplo, la información sobre la tolerancia ajena llega por una plataforma, que decide el alcance). Ninguna fuente verificada separa las contribuciones.

10.3 Información sobre lo que los otros toleran: el experimento de Pittsburgh

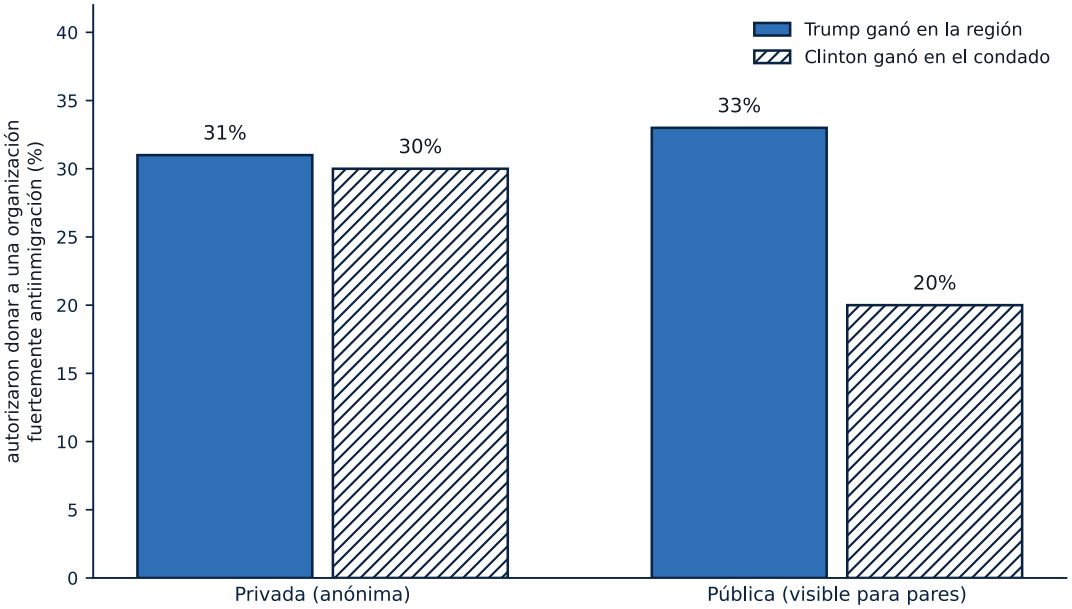
El experimento de la apertura es la prueba más directa, entre las fuentes verificadas, de la hipótesis de que la información sobre lo que los otros aceptan desplaza lo que se hace en público. Bursztyn et al. (2020) la realizaron en dos experimentos que miran los dos lados de la interacción: quien actúa y quien observa y puede sancionar. En el primero, de agosto a octubre de 2018, una empresa de panel en línea reclutó a 1.600 participantes de los siete condados del área metropolitana de Pittsburgh. Al azar, a la mitad se le informó que

Trump había ganado en el área metropolitana en la elección de 2016, y a la otra mitad, que Clinton había ganado en el condado de Allegheny, donde está Pittsburgh. Se ofreció, entonces, una bonificación en dinero a quien autorizara una donación, en su nombre, a una organización fuertemente antiinmigración (para los autores, aceptar era una acción xenófoba lucrativa), y a la mitad de los participantes se les aseguró el anonimato (condición privada), mientras que a la otra se le hizo creer que la decisión sería visible para pares de la región (condición pública).

En privado, las tasas de donación fueron del 31% (ganó Trump) y del 30% (ganó Clinton), sin diferencia significativa ($p = 0{,}894$); en público, del 33% y del 20% ($p < 0{,}001$). La diferencia entre las diferencias, esto es, cuánto el efecto de la información en público excede el efecto en privado, fue significativa ($p = 0{,}008$). La percepción media de la popularidad local de las visiones antiinmigración fue del 50,1% entre quienes supieron que Trump había ganado y del 42,6% entre quienes supieron que Clinton había ganado. En la estimación por variable instrumental, que usa el sorteo de la información para estimar el efecto de la percepción sobre la donación, un punto porcentual más en esa percepción aumentó en 2,3 puntos porcentuales la donación en público y no alteró la privada, resultado que los propios autores tratan como sugerente, porque el tratamiento puede actuar por otros canales, aunque los consideran improbables.

FIGURA 10.2

Donación a una organización antiinmigración, en privado y en público, según quién ganó la elección en el área



N = 1.600, área metropolitana de Pittsburgh. Diferencia entre las diferencias: p = 0,008.

Nota. Valores redondeados, como en el artículo (Bursztyn et al., 2020). N = 1.600, área metropolitana de Pittsburgh, 2018; la información sobre el ganador se sorteó. El tratamiento es la información sobre quién ganó la elección en el área, y no un dicho de un candidato.

La Figura 10.2 muestra dónde está la diferencia, y es más fina que "la victoria de Trump liberó la xenofobia". Los autores observan que el resultado viene sobre todo de la brecha entre donar en privado y en público en la condición Clinton (30% frente a 20%), sin diferencia significativa en la condición Trump (31% y 33%; $p = 0{,}594$). Lo leen como indicación de que, en Pittsburgh, la victoria de Trump no generó una presión significativa para que personas tolerantes actuaran como xenófobas en público, mientras que saber de la victoria de Clinton se asoció a la presión social para que los xenófobos parecieran tolerantes en público. La lectura que hacen es que la información no vuelve más xenófobos a los participantes: deja más a gusto, para expresar sus visiones en público, a los que ya lo eran.

El segundo experimento mira el lado de quien observa y puede sancionar. En noviembre de 2018, cerca de 1.800 encuestados que se identificaban como demócratas dividieron US$ 2 con otra persona en un juego del dictador, en el que quien decide cuánto transferir se queda con el resto. La otra persona era un participante del primer experimento que había hecho la donación a la organización antiinmigración (solo entraron los que donaron), y el encuestado no sabía que el área era Pittsburgh. Dos datos sobre ese donante se cruzaron al azar: quién había ganado la elección en su área, Trump o Clinton, y si la donación había sido privada o visible para pares, lo que da cuatro condiciones. El valor medio dado fue de US$ 0,78, US$ 0,81 y US$ 0,78 en las tres condiciones en las que no había razón para suponer presión para actuar como xenófobo (las dos privadas y la pública donde ganó Clinton) y de US$ 0,88 en la condición de donación pública en un área donde ganó Trump; la proporción que no dio nada fue del 18%, 21% y 22% en las tres y del 8% en la última. Los autores explican el resultado por el hecho de que un donante público en un área donde venció Trump puede estar bajo presión social para actuar como xenófobo, lo que debilita la inferencia de que sea realmente xenófobo; por eso recibe más, esto es, es menos castigado.

En el lenguaje aquí usado, el experimento 1 actúa sobre el borde del lado del hablante y el 2, sobre el de la sanción del oyente, y los dos bordes se mueven con información del mismo tipo, la de quién ganó la elección en el área. La información no cambió lo que los participantes aceptaban hacer donde nadie vería, y cambió, o dejó de estrechar, lo que aceptaban hacer donde otros verían. Eso es compatible con la definición de la sección 10.1, en la que la ventana es propiedad de la relación con una audiencia, y no del hablante. Los auto-

res adoptan una visión estrecha de norma social, en la que esta guía acciones públicas o potencialmente públicas y no privadas, y dicen que lo que manipularon fue la popularidad percibida de Trump, y no dichos suyos.

Los límites son nítidos. La variable manipulada es la información sobre quién ganó la elección en el área, y no lo que dijo un político; el desenlace es autorizar una donación, y no proferir un enunciado; la muestra es de un panel en línea de una región y, en el experimento 2, solo de demócratas; y el caso es uno solo, de un lado de la cuestión. Lo que el experimento sostiene es que la información sobre el apoyo local a una posición altera lo que las personas aceptan hacer en público respecto de ella, sin alterar lo que hacen en privado. Que lo mismo valga para dichos, para otras posiciones y para otras audiencias es una extensión aquí propuesta.

DATO	Quienes supieron que Trump había ganado en el área aceptaron donar en público a una organización antiinmigración en una proporción semejante a la de quienes donarían en privado (33% y 31%); quienes supieron que Clinton había ganado aceptaron menos en público que en privado (20% y 30%). Los observadores demócratas dieron menos a donantes públicos en un área donde ganó Clinton que en un área donde ganó Trump (Bursztyn et al., 2020).
INTERPRETACIÓN	La información sobre la popularidad local de una posición altera el costo percibido de expresarla en público, para quien actúa y para quien sanciona, sin alterar la posición privada. En los términos aquí usados: se desplaza la ventana de la audiencia, no la disposición del hablante.
HIPÓTESIS	Lo mismo vale para enunciados, para otras posiciones y para otras audiencias. Los autores no manipulan dichos.

10.4 Dos elecciones, dos medidas: Brasil, 2018, y Estados Unidos, 2016

Si la información sobre lo que los otros toleran desplaza lo que se hace en público, una elección sorprendente es una candidata natural a experimento de la naturaleza: informa, de una vez y a todos, que el apoyo a un candidato y a lo que dice es mayor o menor de lo imaginado. Dos estudios verificados intentan leer eso en texto de redes sociales y llegan a resultados que no se reducen uno al otro.

Marino-Fages y Martínez-Pozo (2026) estudian la elección presidencial brasileña de 2018. Tratan el resultado como sorpresa: Bolsonaro obtuvo el 46% de los votos en la primera vuelta, muy por encima de las encuestas, que preveían cerca del 35%, y la sorpresa varió entre municipios, porque el apoyo a él varió mucho de un lugar a otro. Recogieron más de 37,6 millones de tweets en portugués, geolocalizados en Brasil, sin retuits, publicados los lunes de julio de 2017 a diciembre de 2019, y atribuyeron cada tweet a un municipio por la latitud y la longitud. Para medir el discurso de odio usaron dos métodos. El primero fue un modelo de lenguaje en portugués, BERTimbau (una versión de BERT), ajustado con 5.668 tweets de 2017 que otros investigadores habían etiquetado como con o sin discurso de odio. El segundo fue un diccionario de términos que clasifica el blanco del odio en seis categorías (odio político, blanco político, homofobia, racismo, sexismo e insultos). El desenlace en cada municipio es la proporción de tweets clasificados como odio.

El diseño es de diferencias en diferencias: se compara el cambio de la proporción de odio entre antes y después de la elección, excluido el período de agosto a octubre de 2018, y se ve cómo ese cambio varía con la proporción de votos de Bolsonaro en la primera vuelta en cada municipio. La identificación supone que, sin la elección, las tendencias habrían sido paralelas entre municipios con diferentes niveles de apoyo a él. En el conjunto del país, la proporción media de tweets de odio según BERT subió del 8% (julio de 2017 a julio de 2018) al 9% (2019), cerca de 3.000 tweets más en un día medio.

El resultado central es que el discurso de odio aumentó después de la elección y aumentó más donde Bolsonaro obtuvo menos votos: en municipios con una proporción de votos una desviación estándar mayor, el aumento fue 0,033 desviaciones estándar menor con el clasificador BERT y 0,018 con el diccionario (la desviación estándar, medida de la variación típica de cada variable, pone las dos en la misma escala). El aumento se concentró en homofobia y sexismo, temas de la retórica del candidato; para insultos generales, odio político y blancos políticos, no hubo un patrón comparable, y los autores concluyen que el resultado no está impulsado por una polarización creciente. El número de usuarios que pasaron a publicar contenido de odio sin haberlo hecho antes (el margen extensivo) creció más donde él era menos popular, y la intensidad de la expresión de quienes ya habían entrado (el margen intensivo) creció más donde tenía más apoyo. La proporción de usuarios que publican discurso de odio subió un 10,1% entre hombres, un 6,1% entre mujeres y un 3,3% entre cuentas seudónimas o no identificables. Como ejemplo numérico, los autores toman Río de Janeiro (58,3% de los votos para Bolsonaro; cerca de 32 mil tweets por día hábil, 9% con discurso de odio): un municipio semejante, con

una proporción de votos 10 puntos porcentuales menor, generaría de 40 a 50 tweets de odio más por día, lo que llaman modesto en el nivel individual, pero acumulativo y de amplia participación.

La interpretación de los autores es que el patrón es compatible con una ruptura de la espiral del silencio, en la que la percepción de ser minoría en una cuestión de valores lleva al autosilenciamiento por miedo al aislamiento social: la elección reveló que las visiones asociadas a la retórica del candidato estaban más difundidas de lo que se creía, lo que redujo el costo social percibido de expresarlas, sobre todo donde él era menos popular. Es, en el vocabulario aquí usado, la tercera candidata de la sección 10.2, y coincide con la lectura de Bursztyn et al. (2020). A favor del diseño, los autores observan que Brasil no tenía, en el período, un marco legislativo amplio sobre moderación de contenido en plataformas, lo que vuelve improbable que nuevos regímenes legales expliquen el aumento. Caben cuatro salvedades. El estudio es un documento de trabajo, y la propia portada dice que no pasó por revisión de pares. El clasificador tuvo exactitud, precisión, exhaustividad y F1 de 77 a 79% en los conjuntos de validación y de prueba (la exactitud es la proporción de tweets clasificados correctamente; la precisión, la de aciertos entre los marcados como odio; la exhaustividad, la de los tweets de odio que encontró; el F1 combina las dos últimas). Los tweets de los lunes y los geolocalizados se tratan como muestras representativas del universo de tweets, supuesto de los autores, y la razón de geolocalizados respecto del total varía del 4% al 8% por palabra. Y el diseño mide texto, y no percepción de costo social: el mecanismo es lectura de los autores, y no resultado.

Siegel et al. (2021) estudian la elección estadounidense de 2016. Reunieron dos conjuntos de tweets, del 17 de junio de 2015, el día siguiente al del anuncio de la candidatura de Trump, al 15 de junio de 2017: todos los tweets que mencionaban a Trump o a Clinton (más de 600 millones sobre él y más de 150 millones sobre ella), recogidos en tiempo real, de modo que los tweets después borrados o de cuentas después suspendidas siguen en los recuentos; y casi 400 millones de tweets de una muestra aleatoria de 500 mil usuarios estadounidenses, para ver si el odio también aumentó fuera de la conversación política.

Definieron el discurso de odio como lenguaje hostil y malicioso, motivado por un sesgo, dirigido a una persona o grupo por características innatas, reales o percibidas; el elogio al Ku Klux Klan o el orgullo excesivo del propio grupo, en sí, no cuentan, y la retórica nacionalista blanca se midió aparte. Usaron dos métodos. El primero fue un diccionario de 3.268 términos (insultos contra asiáticos, negros, inmigrantes, musulmanes, judíos, homosexuales y mujeres, más la retórica nacionalista blanca), tomado de bases públicas. Como muchos términos aparecían en nombres de usuario, dentro de otras palabras o en

tweets que condenaban el odio, cerca de 25.000 tweets con esos términos se clasificaron a mano, por tres evaluadores cada uno, y sirvieron para entrenar clasificadores (Naive Bayes) que eliminan esos falsos positivos; el de discurso de odio tuvo un 94% de exactitud, un 95% de precisión y un 90% de exhaustividad. El segundo método prescinde de saber las palabras de antemano: mide cuánto se parecen los tweets de cada día a comentarios de comunidades de Reddit conocidas por contenido racista y de derecha alternativa, tomados como discurso de odio "en estado natural", y solo las comparaciones relativas, como la de los tweets posteriores a la elección con los del inicio de la campaña, son interpretables.

El análisis es de serie temporal interrumpida, que compara el nivel diario antes y después de un evento, y los autores calcularon cuatro medidas (número y proporción de tweets y de usuarios con odio por día), que cuentan esencialmente la misma historia. En cualquier mes, entre el 0,1% y el 0,3% de los tweets contenían odio. No observaron un aumento persistente de discurso de odio ni de lenguaje nacionalista blanco ni a lo largo de la campaña ni después de la elección de Trump; los eventos clave generaron picos breves, que se disiparon rápidamente, y el mayor pico mensual en los datos sobre Trump, a fines de enero de 2017, fue en gran parte de odio misógino después del decreto de restricción de viajes. Dicen que el diccionario es conservador, porque exige que el tweet contenga un insulto o un término despectivo. Limitan el alcance a Twitter, observan que el discurso de odio puede haber proliferado en plataformas especializadas y dejan para investigaciones futuras el efecto de la suspensión de cuentas, y advierten que ver cierto discurso en un momento no significa que haya aumentado, dado el carácter en ráfagas de Twitter, en el que los temas ganan protagonismo por poco tiempo y vuelven al equilibrio.

Los dos estudios no se contradicen de modo simple, y tampoco se suman. Difieren en la medida (un clasificador entrenado con tweets etiquetados y un diccionario por blanco, frente a un diccionario que exige insulto y una clasificación por comunidades), en el diseño (diferencias en diferencias entre municipios, frente a series temporales de la campaña y de la elección), en el país, en la plataforma y en la pregunta: el aumento diferencial según la popularidad local del candidato, por un lado, el nivel agregado del discurso en el Twitter estadounidense, por otro. Un contraste local puede detectar lo que una serie agregada no detecta, y lo contrario también ocurre; a qué se debe la diferencia es lo que los datos no deciden. El desacuerdo es parte del dato, y el capítulo no lo resuelve.

DATO	En Brasil, el discurso de odio en los tweets geolocalizados aumentó después de la elección de 2018, más donde Bolsonaro obtuvo menos votos, concentrado en homofobia y sexismo (Marino-Fages y Martínez-Pozo, 2026, documento de trabajo). En Estados Unidos, no hubo un aumento persistente de discurso de odio en Twitter durante la campaña de 2016 ni después de la elección, con la medida adoptada (Siegel et al., 2021).
INTERPRETACIÓN	Los resultados son compatibles con la idea de que una elección informa sobre la tolerancia ajena, y discrepan en cuanto al efecto agregado; las medidas, los diseños y los contextos difieren demasiado para que uno confirme o refute al otro.
HIPÓTESIS	La información traída por la elección redujo la sanción percibida por hablantes que antes se contenían. Ninguno de los dos estudios mide la sanción percibida ni la que de hecho ocurrió.

10.5 Políticos y redes sociales: quién habla desde arriba y qué recibe

Si, como dice el Mackinac Center (sección 10.1), los políticos son más seguidores que líderes, queda por saber a quién siguen. Frimer et al. (2023) ofrecen el estudio verificado que más se acerca a responder, para políticos estadounidenses en Twitter, y habla menos de "el público" y más del retorno que la plataforma devuelve. Aplicando un clasificador de incivilidad a 1,3 millones de tweets de miembros del Congreso desde 2009, los autores observan un aumento del 23% en la incivilidad a lo largo de una década (de 11,8 en 2009 a 14,5 en 2019). El clasificador, PerspectiveAPI, se hizo para ayudar a la moderación de foros y da a cada texto una nota de 0 a 100, que refleja la probabilidad de que un observador lo considere grosero o irrespetuoso. El aumento se mantuvo en los modelos que dan a cada político su propio nivel de base (intercepto por político), de modo que es más atribuible a políticos que se volvieron más inciviles que a la entrada de políticos más inciviles. En una muestra seudoaleatoria de 998.935 tweets de Estados Unidos, la incivilidad también subió entre civiles (de 21,1 a 27,0, 28%), pero el alza entre los políticos persistió con ese control; subió en el Senado y en la Cámara y en todo el espectro político, con la mayor elevación entre demócratas liberales, diferencia que los autores explican sobre todo por reacciones al presidente Trump. En los tuits de los presidentes, la incivilidad media fue de 13,4 con Obama, sin cambio en el tiempo, y subió a 18,8 en 2017 y a 23,0 en 2019 con Trump.

La parte que interesa al análisis operante es la del retorno. Los tuits extremadamente inciviles recibieron en promedio diez veces más retuits y ocho veces más "me gusta" que los extremadamente civiles, y la razón creció: en 2009, los inciviles recibían cerca de dos veces más retuits; en 2019, casi 15 veces ("me gusta": 1,3 veces y casi 8 veces). El análisis de mediación pregunta si el retorno explica el aumento: al controlar estadísticamente el retorno en "me gusta", el aumento de la incivilidad cayó un 74%; al controlar el de retuits, desapareció. Como la mediación es solo correlacional, los autores la completaron con un análisis de causalidad de Granger dentro de cada cuenta, que compara los momentos en que la incivilidad fue especialmente bien recibida con la incivilidad que vino después, y el análisis mostró que los períodos de mejor recepción fueron seguidos de más incivilidad. Concluyen que los políticos aprendieron de la experiencia positiva de tuitear con incivilidad y, por eso, la intensificaron, y atribuyen el aumento tanto a la sustitución, en el Congreso, de políticos menos inciviles por otros más inciviles como al cambio de los propios políticos.

Dos resultados del mismo estudio enseñan la cautela. El primero: en los discursos en el pleno del Congreso (1995 a 2020) y en los debates presidenciales (1976 a 2016), la incivilidad no subió y puede haber caído, lo que los autores toman como apoyo a la idea de que las redes sociales pueden fomentarla. El segundo: los "me gusta" y los retuits no implican aprobación. En estudios del suplemento, cuantos más "me gusta" recibió un tuit, menos les gustó de hecho a las personas, y los retuits de tuits inciviles parecen impulsados por el valor de entretenimiento, y no por la aprobación moral; el recuento absoluto de "me gusta" está confundido por la mayor exposición de los tuits inciviles, y para corregirlo los autores crearon la razón de "me gusta" por retuit. Registran además que la muestra reúne a 841 de los 1.003 miembros del Congreso de las legislaturas 111 a 116, que la API limita la recolección a 3.240 tweets por cuenta, y que la concordancia del clasificador con observadores humanos fue $r = 0{,}627$.

Leído con el vocabulario del capítulo, el estudio muestra tres cosas. La primera es que lo que el político emite en Twitter acompañó el retorno que la plataforma devuelve, y que los enunciados en el borde (los inciviles) recibieron más retorno que los civiles. La segunda es que la audiencia a la que el político "sigue" no es "el público", sino el subconjunto que reacciona, cuya reacción, según los propios autores, diverge de la aprobación. La tercera es que la incivilidad no subió en las arenas presenciales (pleno, debates), lo que es compatible con la idea de que la ventana es de cada arena, y no del político, como supone la definición de la sección 10.1. El Capítulo 9 (sección 9.6) ya vio, en congresistas estadounidenses, que las publicaciones sobre el exogrupo se

compartieron más en los dos campos (Rathje et al., 2021); el estudio de Frimer et al. (2023) muestra un aumento de la incivilidad en todo el espectro político, con la mayor elevación entre demócratas liberales.

El estudio no responde si los políticos movieron la ventana de la audiencia. Mide la conducta del político y el retorno que recibe, y no lo que los electores pasaron a decir. Los tres diseños reunidos hasta aquí cubren cada uno un lado: Bursztyn et al. (2020) manipulan la información que las personas tienen sobre el apoyo local; Marino-Fages y Martínez-Pozo (2026) observan el discurso de usuarios después de una elección; Frimer et al. (2023) observan el discurso de políticos y el retorno. Ninguno vincula el discurso del político con el de la audiencia mediante un diseño que identifique la dirección. Preguntar si el político lidera o sigue, tal como lo formula la página del Mackinac Center, es preguntar lo que ninguna de las tres fuentes mide.

Hay, por último, una observación de método que sirve de advertencia. Kahmann y Heyer (2019) proponen un método de procesamiento de lenguaje natural para encontrar, en textos de periódico, enunciados que desplazan los bordes de la ventana. La idea es simple: en un período de referencia (8.067 documentos del periódico *taz* sobre refugiados, de 2013 y 2014), se mapean las palabras que suelen aparecer junto a las del tema, el contexto estándar; después, cada frase de 2015 recibe una distancia, de 0 a 1, que crece con la proporción de palabras que nunca aparecieron en ese contexto, y solo se mantienen las frases con un indicador de enunciado, como verbos de decir y comillas, lo que reduce 14.894 frases con palabra objetivo a 3.126. Los propios autores reconocen que el uso de palabras nuevas no siempre refleja una opinión fuera de la ventana, porque muchas coocurrencias reciben una distancia alta solo por no haberse usado antes, y ese no uso se debe a efectos aleatorios, y no necesariamente a ser políticamente incorrecto. Reconocen también que el método no garantiza que toda frase encontrada sea extrema: entre las de mayor distancia aparece una sobre el éxito educativo de los migrantes, que no parece exceder ninguna ventana (lectura aquí propuesta). Lo que el método mide es novedad de contexto, y no sanción; confundir las dos es el error que el capítulo quiere evitar.

DATO

En 1,3 millones de tweets de congresistas estadounidenses, la incivilidad subió un 23% en una década, con un retorno (retuits y “me gusta”) mucho mayor para los tuits inciviles, el control del retorno reduciendo o eliminando el aumento, mediación correlacional y análisis de Granger a favor del reforzamiento, parte del aumento por sustitución de políticos, y sin aumento en los discursos del pleno (Frimer et al., 2023).

INTERPRETACIÓN	Lo que los políticos emiten en la red varía con el retorno que la plataforma devuelve, y los enunciados en el borde fueron reforzados; la arena importa, porque el mismo político no aumentó la incivilidad en el pleno.
HIPÓTESIS	Los políticos movieron la ventana de sus audiencias, o las siguieron. Ninguna fuente verificada mide el discurso de las audiencias en relación con el de los políticos con un diseño que identifique la dirección.

10.6 Sanción, contracontrol y lo que mantiene el borde en su lugar

El Capítulo 5 trató el discurso de odio como mando disfrazado y como control aversivo; este capítulo toma de él solo lo que sirve a la ventana: lo que hace retroceder el borde. El estudio verificado más claro sobre eso es el experimento de Álvarez-Benjumea y Winter (2018). Participantes de una plataforma de crowdsourcing (180, reclutados en Clickworker, residentes en Alemania y a quienes se pagó tres euros) comentaron, en un foro montado por los autores, nueve fotos de cuatro temas controvertidos (feminismo, derechos LGBT, refugiados y multiculturalismo, y pobreza). Estaban obligados a comentar cada discusión y veían solo comentarios elegidos por los autores, tomados de una sesión previa en la que tres evaluadores externos clasificaron 840 comentarios como amistosos, neutros u hostiles. Las discusiones variaban en cuatro condiciones: base (dos comentarios amistosos, dos neutros y dos hostiles), censurada (el contenido de odio anterior borrado), extremadamente censurada (solo comentarios amistosos) y contradiscurso (los comentarios hostiles acompañados de réplicas que destacan la inaceptabilidad del odio).

Los comentarios de los participantes fueron evaluados por tres jueces ciegos a las condiciones, en una escala de 1 (muy amistoso) a 9 (muy hostil), con una concordancia relativamente alta entre ellos (alfa de Krippendorff de 0,69, índice de concordancia en el que 1 es concordancia perfecta); la media de las tres notas formó la puntuación de hostilidad. Los comentarios hostiles (nota 7 o más) fueron 24 (5,88%) en la base, 4 (0,99%) en la condición censurada, 18 (4,44%) en la de contradiscurso y 15 (4,10%) en la extremadamente censurada; solo la reducción de la condición censurada fue significativa. El contradiscurso no tuvo efecto significativo, y los autores sugieren como explicación posible que los contracomentarios, al ser ellos mismos hostiles, pueden añadir ambigüedad sobre la norma al poner en conflicto la información descriptiva y la prescriptiva. Advierten que no tienen datos sobre el equilibrio entre censura y libertad de expresión y que el anonimato del foro impide procesos de

identificación grupal; el estudio no toma posición sobre censurar. El Capítulo 5 (sección 5.5) retoma, con los mismos datos, la conclusión de los autores de que la conducta normativa en línea está motivada más por normas descriptivas (lo que los otros hacen) que por normas prescriptivas (lo que se debe hacer).

En términos de Skinner (1957), el contradiscurso es una sanción sin poder: partió de pares anónimos, sin control sobre los reforzadores del hablante, y nada en el diseño lo convierte en una audiencia negativa "efectiva". Lejos de ser un hallazgo definitivo, se trata de una lectura analítica derivada de la explicación de los autores (según la cual la propia réplica conlleva hostilidad), ya que la investigación no manipuló la autoridad de la fuente sancionadora. El resultado importa, sin embargo, para la práctica: en los datos del experimento, la hostilidad no disminuyó porque alguien dijo que no debía, disminuyó cuando el ambiente exhibió menos de ella. La misma regla debería valer en el sentido contrario, de modo que un ambiente que pasa a exhibir más lo que antes estaba fuera de la ventana la ensancharía; es una hipótesis aquí propuesta, y el experimento no la puso a prueba.

El contracontrol (Delprato, 2002) entra por el mismo camino. La sanción que castiga un enunciado es una contingencia aversiva socialmente mediada; quien la sufre puede escapar de ella o evitarla, y uno de los cuatro repertorios de evitación descritos en la sección 6.4 es el del lugar, la migración a espacios cerrados y sin moderación. En términos de Skinner (1957), es buscar al confidente, la audiencia no punitiva. De ahí la hipótesis, aquí propuesta: la burbuja del Capítulo 6 es una audiencia no punitiva colectiva para ciertos asuntos, que ensancha la ventana efectiva del hablante mientras él permanece en ella, y el dilema de Skinner aparece cuando una publicación alcanza, al mismo tiempo, la burbuja y una audiencia con una ventana más estrecha. Skinner (1957) dice que, mientras las audiencias permanecen separadas, el dilema no aparece. Si las plataformas juntan en una misma emisión audiencias antes separadas, el dilema se instala con más frecuencia; ninguna fuente verificada mide eso.

TABLA 10.1

Lo que cada estudio manipuló u observó, lo que midió como "decir" y lo que no permite concluir

Estudio	Material y diseño	Lo que midió como "decir"	Lo que observaron	Salvedad
Bursztyn et al. (2020)	Dos experimentos en línea en EE. UU., 2018: 1.600 residentes de Pittsburgh; cerca de 1.800 demócratas. Información sorteada sobre quién ganó la elección en el área	Autorizar una donación a una organización antiinmigración, en privado y en público (acción); en el exp. 2, valor dado al donante	En público, 33% (ganó Trump) y 20% (ganó Clinton); en privado, 31% y 30%. Los observadores dieron más al donante público donde ganó Trump	Manipula información, no dichos de políticos; acción, no enunciado; panel en línea; observadores solo demócratas
Marino-Fages y Martínez-Pozo (2026)	Elección de 2018 en Brasil como sorpresa; diferencias en diferencias por municipio; más de 37,6 millones de tweets geolocalizados	Discurso de odio por clasificador BERT (F1 de 77 a 79%) y por diccionario	Aumento después de la elección, mayor donde Bolsonaro obtuvo menos votos; concentrado en homofobia y sexismo	Documento de trabajo, sin revisión de pares; solo del 4 al 8% geolocalizados; tendencias paralelas supuestas
Siegel et al. (2021)	Series de más de 750 millones de tweets de la elección de 2016 y de casi 400 millones de una muestra aleatoria	Diccionario aumentado por aprendizaje automático y clasificación por comunidades de Reddit	Sin aumento persistente de discurso de odio en la campaña ni después de la elección; picos breves	El diccionario exige insulto; solo Twitter; los autores señalan otras plataformas
Frimer et al. (2023)	1,3 millones de tweets de congresistas estadounidenses, 2009–2019; discursos del pleno y debates	Incivilidad por clasificador (PerspectiveAPI; r = 0,627 con observadores)	+23% en una década; mayor retorno para tuits inciviles; sin aumento en el pleno	Manuscrito aceptado; la incivilidad no es discurso de odio; mediación correlacional; parte por sustitución
Álvarez-Benjumea y Winter (2018)	Experimento en línea con 180 participantes de crowdsourcing; foro con cuatro condiciones	Comentarios con nota de hostilidad igual o mayor que 7	5,88% en la base; 0,99% censurada; 4,44% contradiscurso; 4,10% extremadamente censurada	Participantes pagados y anónimos; sin datos sobre el equilibrio con la libertad de expresión

Estudio	Material y diseño	Lo que midió como “decir”	Lo que observaron	Salvedad
Kahmann y Heyer (2019)	Método aplicado a un periódico alemán (taz); 8.067 documentos de referencia (2013 y 2014) y textos de 2015	Novedad de contexto de cada frase	Lista de frases de mayor distancia	Novedad no es inadmisibilidad; los autores admiten que no toda frase encontrada es extrema

Nota. Los números vienen de los artículos. Las páginas de Bursztyn et al. (2020) son las de la American Economic Review y las de Kahmann y Heyer (2019), las de las actas; las demás fuentes no tienen paginación de revista en las versiones leídas. La columna “Lo que midió como ‘decir’” describe lo que cada estudio contó; ninguno midió la sanción que recibió el decir, y solo el de Álvarez-Benjumea y Winter (2018) la hizo variar.

DATO	En un foro experimental, eliminar el contenido de odio anterior redujo los comentarios hostiles del 5,88% al 0,99%, y añadir réplicas que señalan la inaceptabilidad no tuvo efecto significativo (Álvarez-Benjumea y Winter, 2018).
INTERPRETACIÓN	Lo que el ambiente exhibe como hecho por los otros pesa más, en estos datos, que la réplica verbal; la réplica de pares sin poder no tiene el estatus de audiencia negativa efectiva (lectura aquí propuesta, con base en Skinner, 1957).
HIPÓTESIS	La burbuja es una audiencia no punitiva colectiva para ciertos asuntos, y la plataforma, al juntar audiencias, multiplica los dilemas. Ningún estudio verificado las pone a prueba.

10.7 Cómo situar un dicho en relación con la ventana

El capítulo no ofrece un criterio para decir que un dicho está “fuera de la ventana”: eso exigiría medir la sanción, y las fuentes no la miden. Lo que se puede ofrecer es un conjunto de preguntas que impide los errores más comunes de la conversación corriente. El instrumento es para el analista, y no clasifica dichos.

Las preguntas 1 a 4 vienen de las definiciones de las secciones 10.1 y 10.2; las 5 a 8, de los límites de los estudios reunidos en la Tabla 10.1. La pregunta 7 repite el criterio de todo el libro: el mismo método para todos los campos.

Preguntas para situar un dicho en relación con la ventana

1. ¿Cuál es la audiencia y en qué arena ocurrió el dicho (pleno, red, radio, familia)? La misma persona puede tener ventanas diferentes en arenas diferentes.
2. ¿Qué se sabe sobre la sanción que enunciados semejantes recibieron de esa audiencia, por registro (y no por impresión)? Sin registro, el analista debe decir que no lo sabe.
3. ¿Hay audiencias con ventanas diferentes presentes en la misma emisión (dilema)? La plataforma decide cuáles.
4. ¿Qué cambió entre dos momentos: quién escucha, el poder de sanción, la información sobre la tolerancia ajena, lo que el ambiente exhibe, lo que se refuerza en el borde, la repetición o el alcance?
5. ¿Qué se midió como "decir": texto bruto, clasificador, diccionario, acción? ¿Con qué error y con qué definición de "odio" o de "incivilidad"?
6. ¿Contra qué se comparó: el mismo hablante antes y después, u otros hablantes, y con qué criterio de selección?
7. ¿Se aplicó el mismo criterio a dichos de campos políticos opuestos?
8. ¿Qué no distingue la medida (novedad, incivilidad, odio, acción)?

10.8 Interpretaciones rivales y límites

La lectura del capítulo, según la cual la ventana de una audiencia es el conjunto de enunciados que ella no castiga y el borde se desplaza cuando cambia la sanción, tiene rivales. El recuadro de abajo las reúne, una a una, y los límites del capítulo vienen después de él.

Interpretaciones rivales para el mismo conjunto de datos

Cambio de composición, y no de ventana. Frimer et al. (2023) atribuyen el aumento tanto a la sustitución de políticos menos inciviles por otros más inciviles como al cambio de los propios políticos; lo que se ve en el agregado es la suma de los dos, y solo el segundo es un cambio en el discurso de quien ya hablaba.

Cambio de medida. Los estudios usan clasificadores con F1 de 77 a 79% (Marino-Fages y Martínez-Pozo, 2026), diccionarios que exigen insulto (Siegel et al., 2021) y un clasificador de incivilidad con r = 0,627 con observadores (Frimer et al., 2023); "discurso de odio" e "incivilidad" no son el mismo constructo.

Otros canales, y no la norma. Bursztyn et al. (2020) dicen que el efecto del tratamiento puede actuar por otros canales, aunque los consideran improbables; la lectura en términos de norma es de los autores.

Evento, y no ventana. Marino-Fages y Martínez-Pozo (2026) excluyen los meses de agosto a octubre de 2018 y suponen que, sin la elección, las tendencias serían paralelas entre municipios; el supuesto puede fallar, y otros eventos del período pueden haber actuado de modo diferencial entre municipios (observación hecha aquí).

Metáfora que no se refuta. Si cualquier cambio de discurso vale como desplazamiento de la ventana, el concepto no separa hipótesis. Por eso aquí se lo define por la sanción de la audiencia (sección 10.1), y se pide la medida de la sanción (sección 10.7).

Dirección y simetría. Los estudios de elección tratan de las victorias de Trump y de Bolsonaro; ninguno trata del caso especular. El concepto es simétrico por definición, y el dato más cercano a la simetría, en el sentido de que el aumento fue general, viene de Frimer et al. (2023), donde la mayor elevación ocurrió entre demócratas liberales, explicada por los autores sobre todo por reacciones a Trump; la simetría del efecto de una elección no se ha puesto a prueba.

Fuente interesada. La descripción del concepto viene de un centro que lo usa para promover sus propias políticas (Mackinac Center, 2019).

Muestra y plataforma. Panel en línea de una región y observadores demócratas (Bursztyn et al., 2020), participantes pagados de crowdsourcing (Álvarez-Benjumea y Winter, 2018), Twitter (los demás). Nada de esto dice nada sobre WhatsApp, YouTube o TikTok.

Cinco límites valen para todo el capítulo. Primero, el concepto de ventana, en su origen, trata de políticas públicas y viene de una parte interesada; la traducción a enunciados y audiencias se propone aquí, y ninguna fuente la pone a prueba como tal. Segundo, las medidas son diferentes (percepción, acción pública, texto clasificado) y no se suman; las comparaciones de la Tabla 10.1 son de orden de magnitud y de diseño, y no de efectos. Tercero, ninguna fuente mide la sanción, que es el término central de la definición: el experimento de Bursztyn et al. (2020) toca el lado del observador, y el de Álvarez-Benjumea y Winter (2018) hace variar lo que el ambiente exhibe, pero ninguno mide el castigo que el hablante de hecho sufre. Cuarto, la fuente brasileña es un documento de trabajo sobre una plataforma y una elección, y ninguna mide el discurso de políticos brasileños; esa es la pregunta que se quiere hacer, y las fuentes verificadas no la responden. Quinto, los datos que vincularían el discurso del político con el de la audiencia, con identificación de la dirección, probablemente exigen registros que solo tienen las plataformas, como en los límites del Capítulo 8.

Para quien quiera usar el capítulo en el análisis de dichos (Capítulos 3, 5 y 7), la consecuencia es práctica. Decir que un dicho "desplazó la ventana" es afirmar tres cosas, todas empíricas: que la audiencia lo habría castigado antes, que dejó de castigarlo y que el dicho del político tuvo parte en eso. El analista

que afirma solo la primera cosa describe una impresión. El Capítulo 11 retoma, al tratar de las medidas de difusión, la distinción entre alcance, exposición y reacción, que importa aquí porque la ventana efectiva de una publicación depende de quién la ve; el Capítulo 12 (sección 12.6) trata de intervenciones sobre las contingencias de la plataforma, y el criterio de este capítulo, el mismo para lo que ensancha y para lo que estrecha la ventana, vale para ellas.

Síntesis

La ventana de Overton, en su origen, es un modelo sobre qué políticas puede apoyar un político sin perder apoyo popular, formulado por un centro que lo usa para promover sus propias políticas y que afirma que los políticos son más seguidores que líderes. Aquí se la traduce al análisis de la conducta verbal: la ventana de una audiencia, para un hablante y un asunto, es el conjunto de enunciados que esa audiencia no castiga; es relación, y no propiedad del enunciado; hay tantas ventanas como audiencias, y el dilema de Skinner aparece cuando están presentes dos audiencias con ventanas diferentes (secciones 10.1 y 10.2).

Los datos verificados muestran pedazos diferentes del fenómeno. La información sobre quién ganó una elección local alteró lo que residentes de Pittsburgh aceptaron hacer en público, sin alterar lo que hicieron en privado, y alteró lo que los observadores dieron a quien actuó en público (10.3). Una elección brasileña sorprendente fue seguida de un aumento diferencial del discurso de odio en tweets, mayor donde el candidato obtuvo menos votos, en un documento de trabajo; una serie estadounidense no encontró un aumento persistente en Twitter, con otra medida (10.4). Los congresistas estadounidenses se volvieron más inciviles en Twitter a medida que la incivilidad recibía más retorno, y no en el pleno, con parte del aumento por sustitución de políticos y con la salvedad de que los “me gusta” no son aprobación (10.5). En un foro experimental, eliminar el odio redujo el odio, y la réplica que lo condena no (10.6).

Lo que los datos no muestran es el término central y la dirección que pide la pregunta de fondo. Ningún estudio mide la sanción, ni vincula el discurso del político con el de la audiencia con identificación de la dirección, y ninguno trata de un caso especular. La ventana es, en las fuentes, el nombre de un resultado (el cambio en lo que se hace o se dice en público) y no de un mecanismo; el capítulo ofrece, en la sección 10.7, preguntas para que el analista distinga lo que sabe de lo que supone. Los Capítulos 3, 5 y 7 usan la idea de que la ventana es por audiencia y por arena; el Capítulo 6, la de que la burbuja es una audiencia no punitiva (hipótesis); el Capítulo 8, la de que el alcance decide

cuántos oyen; el Capítulo 9, la de que el político es un productor que responde al retorno; el Capítulo 12, la de que intervenir es modificar las contingencias que mantienen el borde.

Referencias

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), Artículo 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.

instituto WALDEN4 | ACADEMIA WALDEN4

RÁDIO IW4

Artículos, conferencias y clases para escuchar

Lecturas de artículos científicos, conferencias y clases en audio, 24 horas al día, 7 días a la semana: para estudiar en el camino, en el tráfico o en el gimnasio.

24h al día, 7 días a la semana

Acceda, gratis y sin registro

walden4.com.br/radio

CAPÍTULO

11

Infodemiología: cómo se propaga la información y qué se mide

Cinelli et al. (2020), en un estudio sobre la infodemia de covid-19, abren el artículo con un caso. Según ellos, una emisora de noticias anticipó el rumor de que Lombardía, en el norte de Italia, sería cerrada y lo publicó horas antes de la comunicación oficial del primer ministro; como consecuencia, las personas abarrotaron trenes y aeropuertos para ir al sur antes del cierre, lo que, en palabras de los autores, entorpeció la iniciativa del gobierno de contener la epidemia y posiblemente aumentó el contagio. El caso se relata como los autores lo cuentan; la noticia no se leyó, y no se sabe cómo se midió la conducta de las personas.

El caso le sirve al capítulo por una razón que no es la de los autores. La información que circuló no se presenta como falsa: era un rumor sobre una medida que llegó a tomarse, y el daño que se le atribuye viene de cuándo y de cómo las personas actuaron a partir de ella. La salud pública distingue el rumor de la desinformación precisamente por eso, porque el rumor puede resultar verdadero (Calleja et al., 2021). La lectura se propone aquí, y de ella sale la primera lección del capítulo: lo que preocupa a quien estudia una infodemia no es solo la información falsa, es el volumen, la velocidad y la calidad de la información, y la veracidad de cada ítem es una variable entre otras que es importante considerar.

El capítulo pregunta cómo se mide la difusión de información, qué muestran los datos verificados sobre lo verdadero y lo falso, y cómo se relaciona esto con las contingencias de los capítulos anteriores. La sección 11.1 presenta la infodemiología y la palabra infodemia; la 11.2 trata de las medidas de difusión y de la distinción entre alcance, exposición y reacción; la 11.3 examina el re-

sultado de que lo falso se propaga más rápido que lo verdadero; la 11.4, los modelos epidemiológicos aplicados a la información; la 11.5, la infodemia de covid-19, en el mundo y en Brasil; la 11.6, lo que el campo propone hacer; y la 11.7 reúne preguntas para leer un estudio de difusión.

Tres cautelas orientan la lectura. La primera es que el lenguaje epidemiológico es una analogía. Ningún agente infeccioso pasa de una persona a otra; lo que se propaga son enunciados, y cada paso es una respuesta verbal de alguien (Capítulo 3), mantenida por consecuencias que el modelo de contagio no puede representar. La analogía es útil para medir y arriesgada para explicar. La segunda es que las medidas de difusión cuentan respuestas de reenviar, y no creencia, ni efecto sobre lo que las personas hacen después. La tercera es la cobertura: casi todo lo que hay viene de Twitter, en inglés, y del comienzo de la pandemia; para Brasil, el capítulo tiene un único estudio de enlaces sobre vacunas.

11.1 Infodemiología e infovigilancia

El término infodemiología es de Eysenbach, que la define como la ciencia de la distribución y de los determinantes de la información en un medio electrónico, con el fin de informar a la salud pública y a las políticas públicas (Eysenbach, 2009). El nombre une información y epidemiología, y la analogía es literal en la propuesta: así como la epidemiología cuenta casos y busca determinantes, la infodemiología cuenta ocurrencias de información y busca lo que las determina. El artículo de 2009 distingue dos tipos de indicador: los de la oferta, que cuentan lo que se publica (páginas, mensajes, blogs), y los de la demanda, que cuentan lo que las personas buscan y por dónde navegan. Las métricas básicas son la prevalencia de información, el número absoluto o relativo de ocurrencias de una palabra clave o concepto en un conjunto; la incidencia, que es cuántas unidades nuevas de información surgen por unidad de tiempo; y la razón de ocurrencia, que compara un concepto con un término de control, porque el total del conjunto no suele conocerse. Usar esos datos para vigilancia es lo que el autor llama infovigilancia, y a los métodos que analizan lo que está disponible sin involucrar a las personas los llama pasivos.

El ejemplo fundador es de salud, y no de desinformación. Un estudio de 2006, del propio Eysenbach, mostró una correlación entre búsquedas sobre gripe en Google y casos de gripe en la semana siguiente, en Canadá, y que las búsquedas precedían en una semana a las consultas a médicos centinela (Eysenbach, 2009). Como Google no compartía datos de búsqueda con investigadores externos, el estudio compró un anuncio activado por palabra clave, que devolvía estadísticas de búsqueda y de clics. El detalle importa por dos razones. Mues-

tra que las medidas de difusión dependen de datos que controlan las plataformas, y Lazer et al. (2018) dicen que la cooperación de las plataformas con los investigadores será crucial para entender la escala del problema y la eficacia de las intervenciones. Muestra además lo que son esos datos: registros de lo que las personas hacen, medido como buscar y hacer clic, y no de lo que dicen, lo que el autor presenta como una ventaja sobre las encuestas, por ser directos y estar libres de sesgos de recuerdo y de deseabilidad social (Eysenbach, 2009).

La ventaja tiene un costo que el propio autor registra. La búsqueda aislada es ambigua en cuanto a la intención de quien busca (alguien que busca "resfriado" no necesariamente tiene los síntomas), y los recuentos de resultados de los buscadores no son exactos ni confiables (Eysenbach, 2009). En el lenguaje del Capítulo 3, la forma de la respuesta, la palabra buscada, no dice la función. El autor admite además la flecha en el otro sentido: los cambios en los patrones de información pueden afectar la salud de la población, como en un "brote" de desinformación o en una campaña, y en 2009 escribe que no existían metodologías sólidas para identificar y seguir automáticamente la prevalencia de la información que lleva a menores tasas de vacunación (Eysenbach, 2009). La frase es de 2009, y no se afirma que siga siendo verdadera; marca el punto de partida del campo.

La palabra infodemia se hizo conocida en 2020. El 15 de febrero de ese año, el director general de la Organización Mundial de la Salud (OMS) planteó la preocupación de que la epidemia venía acompañada de una infodemia (Calleja et al., 2021), y la organización la define como un exceso de información de calidad variable, en entornos digitales y físicos, durante un evento agudo de salud, que lleva a confusión, a conductas de riesgo y a la pérdida de confianza en las autoridades sanitarias. Eysenbach (2020) escribe que acuñó infodemiología en 2002 y, a lo largo de veinte años, infodemia e infovigilancia; el artículo de 2009, que fue una de las bases consultadas, usa los términos infodemiología e infovigilancia, y no trae la palabra infodemia.

Zielinski (2021) reduce las características primarias de una infodemia a dos, el volumen de información generado y la velocidad con que aparece, y trata las formas de información errónea como características secundarias, aunque sean las más importantes. El autor se pregunta si un gran volumen de información en gran parte verdadera y útil, que surge deprisa, sería una infodemia, y responde que, para quien gestiona información, cualquier brote de información de gran escala y flujo rápido necesita ser gestionado. El cuadro del artículo asocia al volumen problemas de localizar, almacenar, evaluar la calidad, dar visibilidad y validar, y a la velocidad, problemas de evaluar, filtrar, aplicar, rastrear la historia y evitar el desperdicio, como la investigación repe-

tida sin necesidad. Los ejemplos vienen sobre todo de la literatura científica: el aumento de más del 400% en los depósitos de preprints en medRxiv (de 586 en las últimas 15 semanas de 2019 a 2.572 en las primeras 15 de 2020), que el artículo cita de otro trabajo.

Esto reorganiza la pregunta. Los Capítulos 1 a 7 trataron de enunciados falsos y de lo que los sostiene. La infodemiología mide un fenómeno de otro orden: el volumen y la velocidad de la información, verdadera o falsa, y lo que eso le hace a la capacidad de quien la recibe de separar una de otra. Lo falso es parte del problema, y no su definición. Eysenbach (2020) añade una cautela que el capítulo adopta. Al inicio de una pandemia, a los "hechos" es mejor llamarlos "BETs" (best evidence at the time), la mejor evidencia en el momento; las recomendaciones basadas en ellos están sujetas a cambios, y combatir la infodemia difundiendo hechos es más fácil de decir que de hacer cuando no está claro cuáles son los hechos. El autor da como ejemplos las recomendaciones sobre mascarillas, el uso de ciertos medicamentos y las reglas de distanciamiento. La consecuencia para cualquier medida de "información falsa" en salud se propone aquí: la etiqueta de falso depende de quién verifica, con qué evidencia y en qué momento, y el Capítulo 2 trató esa dependencia para las agencias de verificación.

11.2 Medidas de difusión: alcance, exposición, reacción y la estructura de la cascada

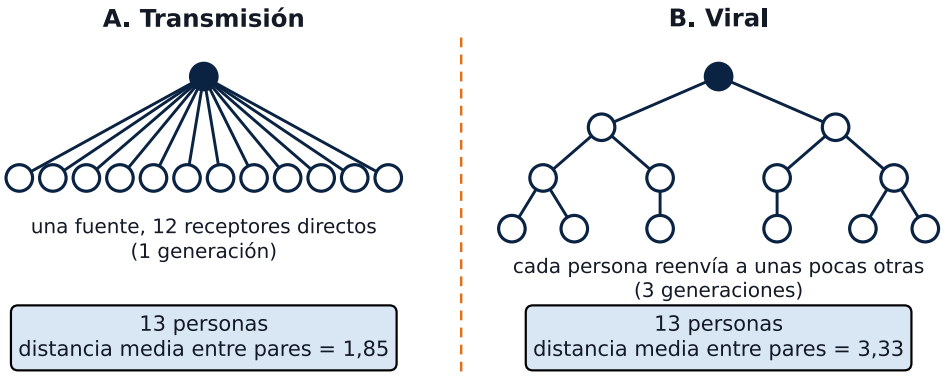
Medir la difusión exige decidir qué cuenta como una persona que "recibió" y qué cuenta como una que "lo pasó a otros". Los estudios del capítulo usan la unidad de la cascada: el conjunto de reenvíos ligados, por una cadena ininterrumpida, a un origen común. Una cascada puede tener tamaño uno, si nadie reenvió la emisión original, y un rumor puede tener varias cascadas, tantas como las veces en que se introdujo de modo independiente (Vosoughi et al., 2018). Cada cascada se describe con cuatro números: el tamaño, que es el número de usuarios involucrados; la profundidad, que es el número de saltos de retuit a partir del origen, donde un salto es un retuit de un usuario nuevo; la amplitud máxima, que es el mayor número de usuarios involucrados en cualquier profundidad; y la viralidad estructural.

La viralidad estructural la utilizan Goel et al. (2016) y responde a un problema de interpretación. Cuando se dice que algo "se viralizó", se entiende que creció de persona a persona, como un virus; pero un contenido puede llegar a millones por una única gran transmisión, como la de una cuenta con decenas de millones de seguidores, sin que nadie lo reenvíe a otro. El tamaño final no distingue los dos casos. La profundidad también falla, porque una única cadena

larga, en una transmisión enorme, da una profundidad alta. Los autores definen la viralidad estructural como la distancia media entre todos los pares de personas del árbol de la cascada, medida que la química matemática conoce como índice de Wiener, y que no depende de supuestos sobre el proceso que generó la difusión. En una transmisión pura, queda cerca de 2, el mínimo; cuantas más generaciones, más alta. La Figura 11.1 muestra dos árboles de 13 personas y la diferencia que el número registra.

FIGURA 11.1

Transmisión y difusión viral: dos árboles del mismo tamaño



Mismo tamaño, estructuras opuestas. Esquema original, a partir del concepto de Goel et al. (2016); la distancia media se calculó para los árboles dibujados. Círculo negro: quien introdujo el contenido.

Nota. Esquema original, a partir del concepto de Goel et al. (2016); sin datos. La distancia media entre todos los pares se calculó para los dos árboles dibujados; en una transmisión con muchas personas, el valor tiende a 2.

El conjunto de Goel et al. reúne cerca de 1.000 millones de eventos de difusión en Twitter, de julio de 2011 a junio de 2012: tuits con enlaces de sitios populares de noticias, videos, imágenes y peticiones. Lo que encontraron deshace la imagen de lo viral. El tamaño medio de los árboles es 1,3, y cerca del 99% de las adopciones son de las raíces o de sus seguidores inmediatos; solo el 0,025% de los árboles llega a 100 nodos o más, 219.855 en cerca de 1.000 millones. Entre ellos hay casi todas las estructuras, de la transmisión pura a las ramificadas, pero la viralidad estructural mediana es baja y varía poco con el tamaño: menos de 3 para imágenes y videos, incluso con 10.000 republicaciones o más; entre 7 y 8 para peticiones; y por debajo de 8 para las mayores cascadas de noticias, de 3.000 republicaciones. La correlación entre tamaño y viralidad estructural es 0,2 para noticias, 0,04 para peticiones y prácticamente cero para imágenes y videos, y los autores concluyen que, si la popularidad se relaciona de modo consistente con alguna característica, es con el tamaño de la mayor transmisión. Sobre las causas dicen que solo pueden especular: sospechan de la falta de grandes canales de transmisión para las peticiones y, para imágenes y videos, de celebridades con decenas de millones de seguidores.

El resultado vale con el límite que ellos mismos marcan: se restringe a la difusión de información en Twitter, y sin reconstruir la secuencia de los eventos y sin una medida de viralidad la mera popularidad permite concluir poco sobre la importancia relativa de los mecanismos viral y de transmisión. También la reconstrucción es una inferencia: cada adoptante se liga al amigo, es decir, a la cuenta que él sigue, que adoptó el contenido más recientemente antes que él, y los autores dicen inferir correctamente el padre en cerca del 95% de los casos.

Tres términos organizan lo que cuentan las medidas, y aquí se fijan porque los Capítulos 8 y 10 los usan y la literatura los mezcla.

DEFINICIÓN PROPUESTA

Alcance, exposición y reacción.

Alcance es el conjunto de personas a quienes se entrega una emisión o que pueden verla; depende de la red de contactos y de la regla de entrega de la plataforma (Capítulo 8). *Exposición* es lo que una persona efectivamente vio. *Reacción* es lo que se hace después de ver: dar "me gusta", comentar, reenviar o ignorar. Skinner no usa los términos; la separación se propone aquí y sirve para decir, de cada medida, cuál de los tres cuenta.

Los estudios del capítulo cuentan sobre todo reacción, y cuando hablan de exposición la infieren. Goel et al. presuponen la exposición por la lista de cuentas seguidas; Mosleh y Rand (2022) construyen la puntuación de exposición de cada usuario a élites que mienten a partir de a quién sigue, y recuerdan que exposición y compartir son constructos fundamentalmente diferentes: la mayor parte de las personas comparte solo una pequeña fracción de lo que ve. El alcance, a su vez, es lo que fija la plataforma, y el Capítulo 8 vio que el ordenamiento altera el alcance de cada emisión de modo desigual (Huszár et al., 2022). La Tabla 11.1 dice, de cada estudio del capítulo, qué cuenta como difusión y qué queda fuera.

Ninguna de esas medidas dice lo que la emisión le hizo a la conducta de quien la vio. Lazer et al. (2018) escriben que exposición no equivale a impacto y que la evaluación del efecto de la exposición a noticias falsas sobre la conducta política a mediano y largo plazo es prácticamente inexistente en la literatura. El capítulo vuelve a este punto en la sección 11.6.

TABLA 11.1

Lo que cada estudio cuenta como difusión y lo que queda fuera

Estudio	Qué cuenta	Unidad	Mide	No mide
Eysenbach (2009)	Búsquedas (demanda) y publicaciones (oferta) con una palabra clave o concepto	Término o concepto en un conjunto de páginas o búsquedas	Ocurrencias y búsquedas	Quién vio; qué hizo con la información; la intención de la búsqueda
Goel et al. (2016)	Republicaciones de enlaces, ligadas a la cuenta seguida que republicó antes	Árbol de cada contenido	Reacción (republicar); la exposición se presupone	Lo que la plataforma entregó; quién vio sin republicar; la veracidad del contenido
Vosoughi et al. (2018)	Retuits de una afirmación verificada	Cascada de rumor	Reacción (retuitear), en número de usuarios	Quién vio sin retuitear; qué rumores nadie verificó
Cinelli et al. (2020)	Personas que publican sobre el tema; publicaciones y reacciones por clase de medio	Autores por día; publicaciones y reacciones por enlace	Reacción (publicar, comentar, reaccionar)	Exposición; la veracidad de cada afirmación (la etiqueta es del medio)
Massarani et al. (2021)	Interacción total del enlace (“me gusta”, comentarios, veces compartido, reacciones)	Enlace entre los cien de mayor interacción	Reacción agregada, sin distinguir el tipo	Alcance; enlaces fuera de los cien; quién reaccionó

Nota. Elaboración propia, a partir de los métodos de los artículos. La columna “Mide” usa los tres términos de la definición anterior.

11.3 ¿Lo falso se propaga más rápido que lo verdadero?

La respuesta más citada del área a esta pregunta es la de Vosoughi et al. (2018), y vale la pena conocer el diseño antes del resultado. Los autores reunieron las cascadas de rumores que seis organizaciones de verificación habían verificado y que circularon en Twitter de 2006 a 2017: cerca de 126.000 historias, tuiteadas por cerca de 3 millones de personas más de 4,5 millones de veces, clasificadas como verdaderas o falsas con un 95 a 98% de concordancia entre las organizaciones. La recolección partió de respuestas, en inglés, a tuits que traían el enlace de una de las páginas de verificación; el tuit original y sus retuits forman la cascada. Los autores evitan la expresión “noticia falsa”, por

considerar que se volvió políticamente polarizada y sin relación con la veracidad. En su lugar, usan "verdadero" y "falso" y dicen que no afirman nada sobre la intención de quien divulga la información.

Lo falso se difundió significativamente más lejos, más rápido, más hondo y más ampliamente que lo verdadero en todas las categorías de información, y el efecto fue más pronunciado para lo falso político que para lo falso sobre terrorismo, desastres naturales, ciencia, leyendas urbanas o finanzas. Lo verdadero rara vez llegó a más de 1.000 personas, y el 1% superior de las cascadas falsas llegaba rutinariamente a entre 1.000 y 100.000; además, lo verdadero tardó cerca de seis veces más que lo falso en llegar a 1.500 personas. En un modelo de la probabilidad de retuitear, lo falso tuvo un 70% más de probabilidad de ser retuiteado, controlando la antigüedad de la cuenta, la actividad y el número de seguidores y seguidos de quien lo originó, y si la cuenta estaba verificada. Un dato pide cuidado: una fracción mayor de rumores falsos tuvo entre 1 y 1.000 cascadas, y una fracción mayor de rumores verdaderos tuvo más de 1.000. Lo verdadero, por lo tanto, tuvo con más frecuencia muchos orígenes independientes, y el resultado de lo falso es un resultado de alcance por cascada.

Los autores proponen una explicación y la presentan con cautela. En todas las métricas de novedad, calculadas frente a los tuits a los que los usuarios habían estado expuestos en los 60 días anteriores, lo falso era más novedoso que lo verdadero, aunque los usuarios pueden no haberlo percibido como novedoso. En las respuestas, según un léxico de emociones, lo falso suscitó más sorpresa y asco, y lo verdadero, más tristeza, anticipación, alegría y confianza. Pero escriben que no pueden afirmar que la novedad cause retuits ni que sea la única razón por la que lo falso se retuitea más, y que la explicación conductual exige investigación con contacto directo con los usuarios. Sobre los bots, los autores encontraron datos de que aceleraron la difusión de noticias verdaderas y falsas al mismo ritmo, y concluyen que lo falso se propaga más porque los humanos, y no los bots, tienen más probabilidad de propagarlo.

Los límites son del diseño. Entran solo rumores que alguien verificó, y las organizaciones pueden elegir ciertos tipos de rumor o llamar la atención sobre ellos; los autores pusieron a prueba la objeción con 13.240 cascadas que estudiantes, trabajando de modo independiente, investigaron y etiquetaron como verdaderas, falsas o mixtas, con un 90% de concordancia y un kappa de Fleiss de 0,88 (índice de concordancia entre evaluadores que descuenta la concordancia esperada por azar), y hallaron resultados casi idénticos. Todo es Twitter, en inglés, hasta 2017. Y "más lejos" quiere decir más usuarios nuevos que retuitearon, y no más personas convencidas. Hay además una tensión aparente con el estudio de Goel et al. de la sección anterior. Vosoughi et al. dicen que

lo falso no se propagó por transmisión simple, sino por difusión de persona a persona, en un proceso de ramificación viral; Goel et al. muestran que, en promedio, las grandes cascadas están dominadas por la transmisión. Las dos cosas caben juntas, porque uno compara lo falso y lo verdadero y el otro describe el conjunto, y recuerdan que "viral" es una medida de estructura, que se compara, y no una etiqueta que se aplica a todo lo que es grande.

Además de Vosoughi et al., otros dos estudios del capítulo hacen la misma pregunta con otra unidad, y la Tabla 11.2 los pone lado a lado.

TABLA 11.2

Tres estudios, tres unidades: lo que cada uno comparó entre lo confiable y lo falso

Estudio	Unidad comparada	Quién clasificó	Plataforma y período	Lo que se halló
Vosoughi et al. (2018)	Afirmación verificada (cascada de rumor)	Seis organizaciones de verificación	Twitter, 2006–2017	Lo falso llegó más lejos, más rápido, más hondo y más ampliamente; el efecto fue mayor para lo político
Cinelli et al. (2020)	Enlace, etiquetado según el medio (confiable o cuestionable)	Media Bias/Fact Check	Cinco plataformas, 1 de enero a 14 de febrero de 2020, sobre covid-19	Las publicaciones y reacciones de fuentes confiables y cuestionables crecen al mismo ritmo en cada plataforma; la amplificación relativa varía con la plataforma
Massarani et al. (2021)	Enlace sobre vacunas, entre los cien de mayor interacción	Los propios autores, con categorías adaptadas de otro trabajo	Facebook, Twitter, Pinterest y Reddit, en portugués, 2018–2019 y 2020	Proporción de desinformación igual en los dos períodos; interacción media mayor para la desinformación solo en 2020

Nota. Los resultados vienen de los artículos; las unidades y los criterios, en las secciones 11.2 y 11.5. No son medidas comparables entre sí.

No se leyó ningún trabajo que reexamine, con otro método, la comparación de Vosoughi et al. Tres trabajos previstos en el plan de este capítulo no pudieron obtenerse por vía legal y automatizada (la lista está en el Apéndice D). El capítulo, por lo tanto, no dice qué concluyen, y el resultado de Vosoughi et al. aparece aquí sin el examen que trabajos posteriores puedan haber hecho de él.

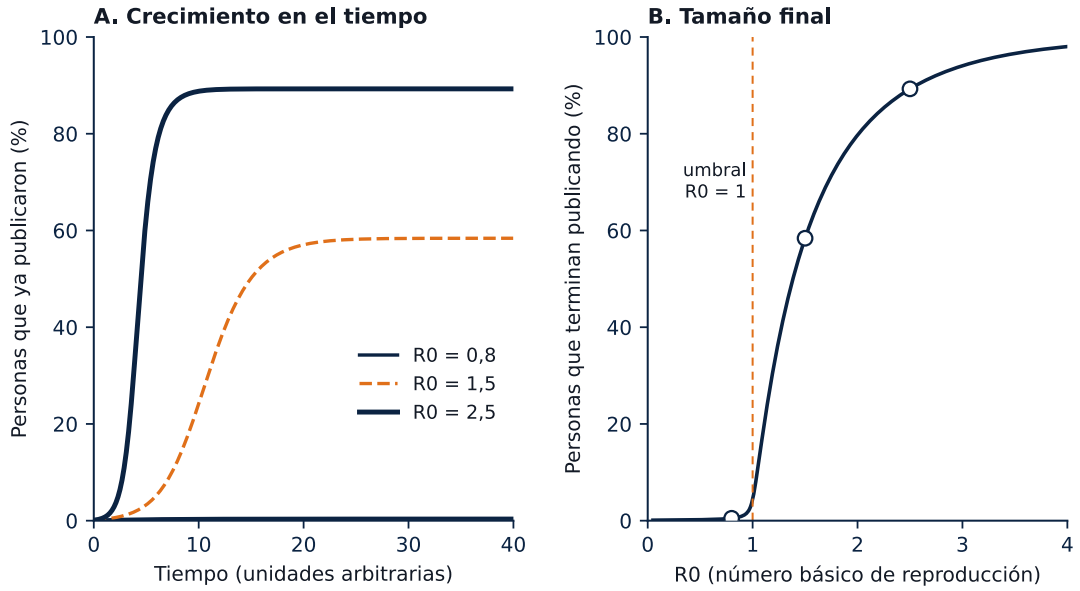
DATO	Afirmaciones verificadas y falsas llegaron más lejos y más rápido que las verdaderas en Twitter, de 2006 a 2017 (Vosoughi et al., 2018). En cinco plataformas y en los primeros 45 días de 2020, los enlaces de medios etiquetados como cuestionables y confiables crecieron al mismo ritmo dentro de cada plataforma, y la amplificación relativa varió entre ellas (Cinelli et al., 2020). En 94 y 89 enlaces sobre vacunas, la proporción de desinformación fue prácticamente la misma en 2018–2019 y 2020, y la interacción media de la desinformación superó a la de lo verificado solo en 2020 (Massarani et al., 2021).
INTERPRETACIÓN	"Lo falso se propaga más" es una afirmación sobre una unidad (la afirmación verificada), una plataforma y un período. La misma pregunta con otra unidad, el medio, o en otra plataforma, no da el mismo resultado, y ninguno de los estudios contradice al otro, porque miden cosas diferentes.
HIPÓTESIS	Cuando lo falso se propaga más, es porque propiedades del contenido, como la novedad, controlan respuestas de reenviar que la audiencia ya emitía, en el sentido de los Capítulos 8 y 9. Es compatible con los datos de novedad, sorpresa y asco de Vosoughi et al., que no tienen prueba causal; los propios autores dicen que no pueden afirmar que la novedad cause retuits.

11.4 Modelos epidemiológicos aplicados a la información

Los modelos epidemiológicos dividen la población en compartimentos, susceptibles, infectados y recuperados (SIR); la persona infectada contagia a cada contacto susceptible con cierta probabilidad; y el número básico de reproducción, R0, es el número medio de casos secundarios que produce un infectado. En epidemiología, R0 = 1 es el umbral: por debajo de él la enfermedad se extingue, por encima se propaga (Cinelli et al., 2020). Goel et al. (2016) escriben R0 como el producto del número medio de contactos de cada nodo por la probabilidad de contagio por contacto. La Figura 11.2 muestra lo que hace el umbral, en una simulación propia, con parámetros ilustrativos y sin datos.

FIGURA 11.2

El umbral de R0 en un modelo SIR simple



Simulación original: modelo SIR de mezcla homogénea (todos tienen el mismo contacto con todos), con un 0,1% de personas iniciales y recuperación de 1 por unidad de tiempo. Parámetros ilustrativos; ningún dato de plataforma, tema o período.

Nota. Simulación original: modelo SIR de mezcla homogénea (todos tienen el mismo contacto con todos), con un 0,1% de personas iniciales y recuperación de 1 por unidad de tiempo. Parámetros ilustrativos; ningún dato de plataforma, tema o período. Panel B: tamaño final calculado para cada R0.

Cinelli et al. (2020) aplicaron el modelo a 1.342.103 publicaciones y 7.465.721 comentarios de 3.734.815 usuarios en cinco plataformas, del 1 de enero al 14 de febrero de 2020. Lo que crece es el número de autores que publican sobre el tema, y el "infectado" es quien pasa a publicar después de estar expuesto a él; en el enfoque de los autores, un R0 mayor que 1 indicaría la emergencia de una infodemia. Ajustaron dos modelos: uno fenomenológico (EXP), que busca reproducir los datos sin suponer un mecanismo de crecimiento específico, y el SIR (*Susceptible*, *Infected*, *Recovered*) estándar. Todos los valores estimados quedan por encima de 1 (son "supercríticos"), incluso considerando los intervalos de confianza: con el EXP, de 1,42 a 1,52 en Gab[1] y de 2,02 a 2,64 en Instagram; con el SIR, de 2,2 a 2,5 en Gab y de 4,0 a 5,1 en Twitter, y cerca de 100 en Instagram, valor que los autores dicen que está muy por encima de lo observado en cualquier epidemia real y que atribuyen a un salto abrupto de nuevos autores que los modelos continuos no explican.

[1] Red estadounidense de microblogs.

Dos cuidados valen para estos números. El primero es lo que R0 es aquí: un parámetro de crecimiento del número de personas que publican sobre un asunto en una plataforma, en 45 días, y no una propiedad de una afirmación falsa ni de una persona. El segundo es que el ajuste del modelo a la curva no establece el mecanismo. Goel et al. (2016) escriben que replicar algunos resultados empíricos con un modelo teórico no implica que este represente el proceso generador verdadero, aunque sea posible descartar algunos modelos.

El uso más instructivo del modelo, en la literatura consultada, es el de Goel et al., que lo emplearon para descartar hipótesis. El modelo en red aleatoria se descarta porque produciría tamaños concentrados en torno a la media, y no la cola pesada; con una red libre de escala y baja infecciosidad (r en torno a 0,5, es decir, subcrítico), el modelo reproduce la probabilidad de que una emisión sea popular, la viralidad estructural media y la correlación con el tamaño, pero no la diversidad de estructuras observada. Los autores concluyen que la noción de punto de inflexión (*tipping point*) es, en gran parte, irrelevante para los eventos virales que estudian, lo que va en contra de la idea común de que la información "se viraliza" al pasar un umbral: todo lo que observaron, incluidos los mayores eventos, cabe en un modelo de baja infecciosidad en el que, en una red con pocos polos de muchos seguidores, un contenido encuentra con facilidad un nodo que lo sostiene.

En lenguaje conductual, lo que el modelo llama infección es una respuesta verbal: publicar o reenviar un mensaje en presencia de la emisión de otra persona. El modelo trata esa respuesta como un evento que la exposición desencadena con probabilidad igual para todos los susceptibles y no tiene término para la consecuencia que la mantiene. Es lo que abrieron los Capítulos 6, 8 y 9: la difusión se refuerza por los efectos sobre las reacciones de una audiencia específica, y la probabilidad de difundir varía según la audiencia (Capítulo 6), el contenido (Capítulo 8) y la historia de quien reenvía (Capítulo 9).

CONCEPTO EN USO

Macrocontingencia (Glenn, 2004), retomada del Capítulo 9

El Capítulo 9 (sección 9.7) distinguió la metacontingencia de la macrocontingencia. La macrocontingencia es la relación entre una práctica cultural y la suma agregada de las consecuencias de la macroconducta que la constituye; esos efectos acumulativos, escribe Glenn (2004), no pueden controlar, como consecuencias operantes, la conducta de cada individuo. Una cascada de 10.000 reenvíos es un efecto agregado de ese tipo: ninguno de los 10.000 se mantuvo por el tamaño de la cascada. Lo que mantiene cada reenvío es el retorno que la persona recibe de su audiencia. Para Glenn, la ingeniería cultural eficaz exige identificar las macrocontingencias que producen efectos indeseables y alterar

las contingencias operantes o las metacontingencias pertinentes. Lectura propuesta aquí: reducir el tamaño de las cascadas pide cambiar lo que mantiene el reenvío de cada uno, y no el tamaño.

Esto no vuelve irrelevante la escala. Lo que preocupa a la salud pública es el efecto agregado, y es en él donde mide el modelo epidemiológico. La distinción es de nivel: el fenómeno se mide en el nivel de la población y se explica, en la lectura propuesta aquí, en el entrelazamiento de las contingencias de diferentes personas.

DATO	Los modelos SIR se ajustan a la curva de crecimiento del número de autores sobre un tema en cinco plataformas (Cinelli et al., 2020), y un SIR en red libre de escala con baja infecciosidad reproduce la distribución de tamaños de las cascadas en Twitter, pero no la diversidad de sus estructuras (Goel et al., 2016).
INTERPRETACIÓN	El modelo describe el crecimiento sin describir lo que mantiene cada respuesta de reenviar; su parámetro de contagio reúne, en un número, alcance y probabilidad de reacción.
HIPÓTESIS	La probabilidad de reenvío es producto de las contingencias del Capítulo 9 (reacción de la audiencia, indignación, pertenencia al grupo) y varía entre personas y contenidos de un modo que el modelo de mezcla homogénea no capta. Es compatible con el fracaso del modelo en reproducir la diversidad estructural, pero ese fracaso tiene otras explicaciones, y ninguna fuente leída pone a prueba la hipótesis.

11.5 La infodemia de covid-19, en el mundo y en Brasil

Dos estudios sobre la infodemia de covid-19, uno a escala mundial y otro sobre Brasil, permiten avanzar en la comprensión de este proceso. El de Cinelli et al. (2020) compara la covid-19 en cinco entornos. Twitter, Instagram y YouTube son plataformas convencionales; los otros dos son Gab, que los autores describen como financiado por donaciones, con muy poco control del contenido y con una base de usuarios considerada de extrema derecha, y Reddit, agregador de noticias con filtrado colectivo. Los autores clasificaron los enlaces como confiables o cuestionables según Media Bias/Fact Check, que etiqueta medios y no artículos, y explicitan ese límite. En cada plataforma, las publicaciones y las reacciones de fuentes cuestionables y confiables crecen al mismo ritmo, y el factor de escala varía fuertemente con la plataforma. En Gab, las publicaciones de fuentes cuestionables eran cerca del 70% de las de fuentes confiables, y las reacciones, cerca de 3 veces mayores. El coeficiente de

amplificación relativa, que es la razón entre las reacciones medias por publicación de fuente cuestionable y de fuente confiable, es cerca de 1 en Twitter, 0,35 en YouTube, 0,55 en Reddit y 3,9 en Gab. Los autores concluyen que la información de fuentes confiables o cuestionables no difiere de modo significativo en los patrones de difusión, y que Gab es el entorno más susceptible a la desinformación. Además, lo que impulsa la difusión son las peculiaridades de cada plataforma y la dinámica de grupo de quienes interactúan con el tema. Lo leemos como una señal de cuán fuerte puede ser el efecto anclado en la historia de quien reenvía; el propio estudio no mide esa historia.

Los límites son los de la recolección. El estudio cubre los primeros 45 días, y el primer dato de Twitter es del 27 de enero, por límites de la interfaz de programación de la plataforma; la recolección por flujo de Twitter es cerca del 1% del volumen, y la de Instagram se hizo a mano. La amplificación es relativa a cada plataforma y depende de la etiqueta del medio, no de la veracidad de cada afirmación. Y los autores señalan, en la introducción, que algunos estudios habían indicado que la información falsa puede propagarse más rápido y más ampliamente, pero que eso podría ser un efecto específico de la plataforma, lo que sus datos no contradicen.

Para Brasil, Massarani et al. (2021) usaron BuzzSumo, herramienta de monitoreo que muestra qué contenidos sobre un tema generaron más interacción en las redes, para recolectar los cien enlaces (el máximo que permite la herramienta) con la palabra “vacina” (vacuna) que tuvieron mayor interacción en 2020. De ellos, 94 enlaces válidos, en portugués - después de excluir cinco enlaces de medios de Portugal y un enlace roto - se compararon con los del 22 de mayo de 2018 al 21 de mayo de 2019, que arrojaron 89 enlaces, con los mismos filtros. Los autores aplicaron la noción de interacción a partir de la suma de veces compartido, comentarios, “me gusta” y reacciones en Facebook, de veces compartido en Twitter, de veces compartido y comentarios en Reddit y de veces compartido en Pinterest, pero advirtieron que la definición se restringió a los protocolos de las redes. Clasificaron la exactitud de cada enlace, comparando la información citada con artículos científicos, fuentes oficiales y otros medios, como verificado, no verificable (como opiniones personales, encuestas y contenido de ficción) o desinformación, esta subdividida en tipos como contenido engañoso, fabricado y conexión falsa, con categorías adaptadas de una propuesta de otros autores. Clasificaron también al emisor de cada enlace: profesional, si tenía una política editorial definida, una institución responsable, autores y equipo identificados y cita de fuentes, o no profesional. La Tabla 11.3 reúne los números.

La media de interacción por enlace aumentó 8,6 veces, de 41.405 a 354.321. La proporción de desinformación prácticamente no cambió, ni la de información verificada. Lo que cambió fue el perfil. En 2018–2019, la desinformación era sobre todo contenido fabricado, es decir, 100% falso y creado para engañar o perjudicar (66,7%), emitido por medios no profesionales (83,3%); en 2020, la mayor parte era conexión falsa (84,6%), y los medios profesionales pasaron a ser los principales emisores (61,5%). La media de interacción de la desinformación, que en 2018–2019 era menor que la de lo verificado (36.200 frente a 40.483), fue mayor en 2020 (412.223 frente a 338.019).

La lectura que hacen los autores es que la mayor interacción con la desinformación sugiere un mayor potencial de viralización en la pandemia, y relacionan la conexión falsa con titulares sensacionalistas y con la dependencia de ingresos publicitarios basados en la interacción, apoyados en otros trabajos. Se registra como interpretación suya: el artículo no mide ingresos ni la edición de los titulares. La lectura converge con la hipótesis del Capítulo 8, de que el retorno por atención selecciona formas de titular, y este estudio no la pone a prueba. Con los porcentajes, 2018–2019 corresponde a cerca de 12 ítems de desinformación y 2020 a cerca de 13, una cuenta reconstruida aquí, pero el artículo se apoya en medidas descriptivas. Diferencias de medias entre 12 y 13 ítems, con la variación de interacción que tienen las redes, no sostienen mucho más que la descripción presentada.

TABLA 11.3

Enlaces sobre vacunas de mayor interacción en las redes, 2018–2019 y 2020

Medida	2018-2019	2020
Enlaces válidos	89	94
Interacción total	3.685.100	33.306.200
Media de interacción por enlace	41.405	354.321
Información verificada	80,9%	81,9%
Desinformación	13,5%	13,8%
No verificable	5,6%	4,3%
Media de interacción: verificada	40.483	338.019
Media de interacción: desinformación	36.200	412.223
Tipo de desinformación: fabricado / conexión falsa	66,7% / 8,3%	15,4% / 84,6%
Principal emisor de la desinformación	Medios no profesionales (83,3%)	Medios profesionales (61,5%)

Nota. Valores del artículo. En 2018-2019 había también un 25% de contenido engañoso entre la desinformación; en 2020, ninguno. “Conexión falsa”: títulos, titulares y elementos visuales que contradicen el contenido (Cuadro 1 del artículo).

Los autores ven, en los dos períodos, una instrumentalización política del debate sobre vacunas, con ejemplos de campos diferentes. En 2018–2019 describen contenido falso y verificaciones sobre la vacuna contra la meningitis y un contenido falso que atribuía a un político extranjero una declaración que él no hizo; en 2020, la Coronavac fue la vacuna más citada en contenido desinformativo (23,1%), aunque no encabezaba las menciones generales, y los autores leen las dudas sobre ella como apoyadas en cuestiones políticas, y no en evidencia científica. La lectura es de los autores: el artículo mide qué vacunas se citaron, y no por qué lo hace quien las cita. Un regreso al Capítulo 1 (sección 1.6) muestra que este reúne los demás datos brasileños sobre desinformación.

DATO	En los primeros 45 días de 2020, el ritmo de crecimiento de publicaciones y reacciones de fuentes cuestionables y confiables fue el mismo dentro de cada plataforma, y la amplificación relativa varió de cerca de 0,35 a 3,9 (Cinelli et al., 2020). En los enlaces sobre vacunas de mayor interacción en Brasil, la interacción media aumentó 8,6 veces entre 2018–2019 y 2020, la proporción de desinformación quedó prácticamente igual, y su perfil cambió, de contenido fabricado por medios no profesionales a titular que contradice el contenido, emitido sobre todo por medios profesionales (Massarani et al., 2021).
INTERPRETACIÓN	En los dos estudios, la infodemia de covid-19 aparece como aumento de volumen y de interacción y cambio de perfil, y no como aumento de la proporción de lo falso. La plataforma y el período cambian el resultado.
HIPÓTESIS	El cambio de perfil, de contenido fabricado por no profesionales a titular engañoso por profesionales, refleja la contingencia de retorno por atención a la que están sujetos los medios profesionales (Capítulo 8). Es la interpretación de los autores del estudio, sin prueba; habría que medir la edición de los titulares antes y después.

11.6 Lo que la infodemiología propone hacer

Eysenbach (2020) propone cuatro pilares para la gestión de la infodemia: monitorear la información (la infovigilancia), construir alfabetización en salud y en ciencia, estimular el refinamiento del conocimiento mediante procesos como la verificación y la revisión por pares, y traducir el conocimiento de modo oportuno y sin distorsiones políticas o comerciales. Describe la crisis en cuatro niveles: ciencia, política y práctica, medios de noticias y redes sociales,

en capas de tamaño proporcional a la cantidad de información que genera cada grupo de actores, y dice que el modelo es opinión de experto, y no el marco de la OMS.

La consulta de la OMS que relata el artículo de Calleja et al. (2021) reunió a 110 participantes de más de 35 países y produjo, mediante juicio estructurado de expertos, una agenda de investigación con cinco ejes, 166 preguntas sugeridas y 49 de alta prioridad. Los ejes son medir y monitorear el impacto; detectar señales y entender la difusión y el riesgo; responder con intervenciones; evaluar las intervenciones y fortalecer la resiliencia; y promover el desarrollo y la aplicación de intervenciones y herramientas. Los autores dicen que los efectos nocivos de la infodemia no pueden gestionarse solo con los enfoques corrientes de comunicación, participación comunitaria y mensajes, y que la respuesta exige intervención en varios niveles: individuo, comunidad, medio, plataforma y política. Escriben que las campañas de comunicación pueden aumentar la visibilidad de los mensajes, pero muchas veces no son eficaces para desmentir afirmaciones falsas, y que la mitad de las organizaciones de verificación no trabaja con profesionales de la salud. Enumeran, entre los desafíos del entorno informativo, la amplificación computacional de mensajes polarizadores, la microsegmentación y el debilitamiento del periodismo local.

Zielinski (2021) añade dos posiciones. La primera es que importa mantener la distinción entre *misinformation* y *disinformation*, porque las soluciones probablemente difieren: grandes dosis de buena información ayudan contra la primera, y la segunda usa los mismos canales de la buena información. Distingue además la 'malinformación', información basada en la realidad usada para infligir daño, definición del Consejo de Europa que el artículo reproduce. La segunda es que las acciones correctivas necesitan salvaguardas de derechos humanos, para no llevar a la censura y a la restricción de la libertad de expresión, y que la tecnología sola no lo resuelve.

Lo que el capítulo lleva a la Parte IV son tres puntos. El primero es la distinción entre las cosas que una intervención puede alterar, y la definición de alcance, exposición y reacción organiza la pregunta: la regla de entrega de la plataforma (Capítulo 12, sección 12.6) altera el alcance; las señales y el entrenamiento alteran la reacción (secciones 12.2 a 12.5); en conjunto, actuar sobre las consecuencias del alcance y de la exposición a la desinformación puede reducir daños, y la reacción puede retroalimentar la propia entrega. El segundo es que la pauta del campo es de preguntas, y no de respuestas: la agenda de la OMS es fruto del juicio de expertos, y aquí se lee como un mapa de lo que se ignora; y lo que se ignora puede involucrar condiciones que orientarían las respuestas hacia un menor grado de desinformación. El tercero es que ninguno de los estudios de difusión leídos mide lo que la información le hizo a la

conducta de salud de las personas, como vacunarse o adherirse a una medida. Miden la difusión, y la difusión es el desenlace que el campo mide con más facilidad. Esto limita lo que se puede decir sobre la eficacia de cualquier intervención evaluada solo por una menor difusión, y el Capítulo 12 retoma esta limitación al evaluar las intervenciones.

11.7 Preguntas para leer un estudio de difusión

Para quien usa el capítulo en la lectura de estudios, de noticias sobre estudios o de informes de plataformas, las preguntas que sugieren las secciones anteriores son las siguientes.

1. ¿Cuál es la unidad que se compara: afirmación verificada, enlace, medio, tema, cuenta?
2. ¿Quién clasificó lo verdadero y lo falso, y con qué criterio (verificador, etiqueta del medio, los propios autores)?
3. ¿Qué se cuenta: alcance, exposición o reacción? ¿Qué se infirió, y a partir de qué?
4. ¿La medida es de nivel (cuánto) o de estructura (cómo, por cuántas generaciones)? El tamaño final dice poco sobre la estructura.
5. ¿Cuál es la base de comparación, la plataforma, el idioma y el período?
6. ¿El estudio mide lo que la información le hizo a la conducta de quien la vio, o solo cuánto se propagó?

Síntesis

La infodemiología es, en la propuesta de Eysenbach, la ciencia de la distribución y de los determinantes de la información en un medio electrónico, con el fin de informar a la salud pública; la infodemia, en el uso de la OMS, es el exceso de información de calidad variable durante un evento agudo de salud. Para Zielinski, la infodemia se define por volumen y velocidad, y lo falso es una característica secundaria. El caso de Lombardía, en el relato de los autores que lo cuentan, muestra que una información no falsa puede tener un efecto dañino por su momento y por lo que las personas hacen con ella (11.1).

Las medidas de difusión cuentan sobre todo reacción (republicar, retuitear, interactuar), infieren la exposición por la lista de a quién se sigue y dejan el alcance a la regla de la plataforma. La viralidad estructural, la distancia media entre los pares de una cascada, muestra que, en el Twitter de 2011–2012, casi todo es pequeño y casi todo es transmisión, y que el tamaño de una cascada dice poco sobre su estructura (11.2). Lo falso llegó más lejos y más rápido que

lo verdadero en Twitter de 2006 a 2017, entre afirmaciones verificadas, y otros dos estudios, con otras unidades, no dan el mismo resultado; las revisiones que preveía el plan no se leyeron (11.3).

Los modelos epidemiológicos describen el crecimiento y permiten descartar hipótesis, como la del punto de inflexión, pero tratan la respuesta de reenviar como un evento desencadenado por la exposición, sin término para la consecuencia que la mantiene; la escala es del nivel de la población, y la explicación, en la lectura propuesta aquí, es de contingencias entre personas (11.4). En la covid-19, en los primeros 45 días de 2020, el ritmo de crecimiento fue el mismo para fuentes confiables y cuestionables dentro de cada plataforma, y en los enlaces sobre vacunas en Brasil la interacción aumentó 8,6 veces, con la proporción de desinformación prácticamente igual y el perfil cambiado (11.5). El campo propone monitorear, alfabetizar, refinar y traducir, en varios niveles, y su agenda es de preguntas; ningún estudio de difusión leído mide el efecto sobre lo que las personas hacen (11.6).

La distinción entre alcance, exposición y reacción, que anunciaron los Capítulos 8 y 10, queda fijada en la sección 11.2 y sirve para leer las medidas de los Capítulos 8 y 9, que cuentan uno de los tres sin separarlos, y para la salvedad del Capítulo 10 de que la ventana efectiva de una publicación depende de quién la ve. El Capítulo 12 parte de lo que aquí quedó abierto: qué se puede cambiar, en cuatro niveles, y con qué evidencia.

Referencias

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), Artículo e30979. https://doi.org/10.2196/30979

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), Artículo 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), Artículo 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), Artículo 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), Artículo 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), Artículo 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), Artículo 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

instituto WALDEN4
walden4.com.br



PARTE IV

Intervenciones

12 Intervenciones: qué se puede cambiar, en cuatro niveles, y con qué evidencia

CAPÍTULO

12

Intervenciones: qué se puede cambiar, en cuatro niveles, y con qué evidencia

Un estudio de campo con anuncios en YouTube sirve de punto de partida. A un video de inoculación, visto durante 30 segundos o más, le siguió, en un plazo de hasta 24 horas, una única pregunta: qué técnica de manipulación usaba una publicación ficticia. La proporción de aciertos fue mayor entre quienes vieron el anuncio que en un grupo de comparación, y los autores describen el resultado como un aumento medio de cerca del 5% en el reconocimiento. El tamaño del efecto, medido por la h de Cohen (una medida de la diferencia entre dos proporciones, en la que 0,2 es, por convención, un efecto pequeño), fue de 0,09, por debajo del menor efecto de interés que los propios autores habían prerregistrado (0,10) (Roozenbeek et al., 2022).

Es uno de los estudios de intervención más cuidadosos entre los leídos: seis experimentos aleatorizados, en los cinco primeros con hipótesis prerregistradas, y un estudio de campo con más de 22 mil respuestas. Es también, por eso, un buen punto de partida. Lo que mide es el reconocimiento de una técnica por parte de quien vio un anuncio. No mide si la persona comparte menos, ni lo que su audiencia hace con lo que ella comparte, ni lo que ocurre una semana después; los autores dicen que el estudio de campo evaluó la susceptibilidad, y no la conducta. Entre lo que el estudio mide y lo que los Capítulos 3 a 11 trataron como aquello que mantiene la declaración falsa, o clasificada por una instancia externa como discurso de odio, hay una distancia, y el capítulo existe para medirla, intervención por intervención.

El análisis hecho en los capítulos anteriores situó el mantenimiento del habla en contingencias, y no en la información: la audiencia que reacciona (Capítulo 9), la regla de entrega que decide el alcance (Capítulo 8), la comunidad verbal

que refuerza y protege al hablante (Capítulo 6) y las prácticas culturales que reúnen todo eso (Capítulos 9 y 10). Cada una sugiere un punto de intervención, y el capítulo las organiza en cuatro niveles, que la agenda de la OMS enumera de modo parecido (individuo, comunidad, medio, plataforma y política; Calleja et al., 2021): el individuo, que es lo que la persona lee, ve o aprende; la interacción y el grupo, que es lo que se hace ante quien habla; la plataforma, que es lo que la regla de entrega y la moderación cambian; y la cultura, que son las prácticas y las metacontingencias.

La Figura 12.1 pone lado a lado once estudios de intervención leídos íntegramente y dice, de cada uno, qué se midió. El recuento es propio, hecho sobre la figura. Ocho estudios midieron juicio o desempeño en una tarea; cinco, conducta observada en una plataforma o foro; cuatro, conducta observada con seguimiento de una semana o más; uno solo, el de Guess et al. (2020), midió las cuatro cosas; y ninguno evaluó una intervención a nivel cultural. La distribución no es casual. El individuo es el nivel en el que resulta barato probar, porque se le muestra un material a alguien y se le pregunta qué le pareció. Los otros niveles exigen acceso a cuentas, a foros y al ordenamiento, y es en ellos donde la medida pasa a ser lo que las personas hacen.

Tres cautelas orientan la lectura. La primera es que aquí no se probó ninguna intervención. Los estudios leídos son de terceros, interpretados a la luz del análisis operante, y cuando se dice que una intervención "enseña a discriminar" o "actúa sobre la contingencia", la frase es una lectura propuesta aquí y viene marcada como hipótesis. Tampoco se leyeron las revisiones y réplicas posteriores sobre la corrección de afirmaciones, ni el estudio de Twitter que describe uno de los artículos citados (sección 12.3); donde eso importa, el texto lo dice. La segunda es la medida. El Capítulo 1 mostró que la respuesta a una encuesta es conducta bajo control de la situación en que se dio, y el Capítulo 11, que ningún estudio de difusión leído mide lo que la información le hizo a la conducta de las personas. Lo mismo vale aquí: juzgar un titular en una pantalla no es compartirlo, y compartirlo no es cambiar lo que se hace; la correspondencia entre decir y hacer, que el Capítulo 7 (sección 7.4) trató como efecto de contingencia, y no como dato, vale para la intención declarada de compartir. La tercera es el criterio. El Capítulo 10 usó el mismo para lo que ensancha y para lo que estrecha la ventana de Overton, y el capítulo lo usa para las intervenciones: la misma pregunta, el mismo tipo de medida y la misma salvedad, cualquiera que sea el campo político de quien emite el contenido que se quiere reducir.

El recorrido sigue los niveles y, dentro del individuo, el momento de la intervención. La sección 12.1 trata de corregir después, con la corrección y la advertencia; la 12.2, de corregir antes, con la inoculación leída como entrena-

miento en discriminación; la 12.3, de los estímulos que señalan precisión. La 12.4 trata de lo que se hace ante quien habla (contradiscurso, compromiso público y la hipótesis de los autoclíticos de fuente); la 12.5, de la reducción del prejuicio mediante equivalencia y mediante cómics (historietas); la 12.6, de las contingencias de la plataforma. La 12.7 reúne lo que aquí se entiende por diseño cultural y los límites éticos de quien diseña.

FIGURA 12.1

Qué midió cada estudio de intervención

	Juicio o desempeño en una tarea	Intención declarada de compartir	Conducta observada (plataforma o foro)	Seguimiento de una semana o más
Individuo: lo que la persona lee, ve o aprende				
Nyhan y Reifler (2010) *corrección dentro de una noticia simulada*	●	–	–	–
Freeze et al. (2021) *advertencia retrospectiva de desinformación*	●	–	–	–
Roozenbeek y van der Linden (2019) *juego de inoculación*	●	–	–	–
Roozenbeek et al. (2022) *videos de inoculación*	●	●	–	–
Guess et al. (2020) *consejos para identificar noticias falsas*	●	●	●	●
Dixon y Lemke (2007) *entrenamiento en equivalencia (adultos)*	●	–	–	–
Mizael et al. (2021b) *entrenamiento en equivalencia (niños)*	●	–	–	●
Interacción y grupo: lo que se hace ante quien habla				
Hangartner et al. (2021) *contradiscurso por empatía, en Twitter*	–	–	●	●
Tsipursky et al. (2018) *compromiso público con la verdad*	–	–	●	●
Plataforma: lo que cambian la regla de entrega y la moderación				
Álvarez-Benjumea y Winter (2018) *eliminación de comentarios hostiles (foro simulado)*	–	–	●	–
Nyhan et al. (2023) *relegar fuentes de la misma inclinación (Facebook)*	●	–	●	●
Cultura: prácticas y metacontingencias				
Ninguna intervención a nivel cultural se evaluó en los estudios leídos para este libro.				

● *la variable se midió (el resultado puede haber sido nulo).* – *no se midió.*
Mapa original, a partir de los estudios leídos (Capítulos 4 y 12). Seguimiento: Guess et al., cerca de 20 días (EE. UU.); Hangartner et al. y Tsipursky et al., cuatro semanas; Nyhan et al., tres meses; Mizael et al., seis semanas. La medida el mismo día o en hasta 24 horas no cuenta como seguimiento.

Nota. Mapa original, a partir de los estudios leídos (Capítulos 4 y 12); no aporta datos nuevos. La marca registra que la variable se midió, y el resultado puede haber sido nulo. Seguimiento: cerca de 20 días en Guess et al. (en los Estados Unidos), cuatro semanas en Hangartner et al. y en Tsipursky et al., tres meses en Nyhan et al. (2023) y seis semanas en Mizael et al. (2021b); la medida el mismo día o en hasta 24 horas no cuenta como seguimiento. Las lecturas de cada celda están en las secciones 12.1 a 12.6.

12.1 Corregir después: la corrección, la advertencia y el "efecto rebote"

La intervención más común ante una afirmación falsa es la corrección: después de que la afirmación circuló, alguien la desmiente. Nyhan y Reifler (2010) la probaron en cuatro experimentos en línea, con estudiantes de grado de una universidad católica del Medio Oeste de los Estados Unidos, en el otoño de 2005 y la primavera de 2006. Los participantes leían una noticia simulada que traía una afirmación engañosa de un político, con o sin corrección, y decían después si estaban de acuerdo con ella. Los autores eligieron un diseño entre sujetos para maximizar el efecto de las correcciones, porque medir antes y después anclaría las respuestas en la primera. Definen las percepciones erróneas como creencias falsas o no comprobadas sobre el mundo, y llaman backfire al caso en que la persona pasa a apoyar la opinión original con más fuerza aún.

Los resultados no son uniformes. En el Estudio 1 (130 participantes), la corrección del informe Duelfer sobre armas de destrucción masiva en Irak no redujo la percepción errónea en el conjunto; entre los conservadores (n = 33), el acuerdo con que Irak tenía tales armas subió del 32% en el control al 64% con la corrección, y entre los no conservadores (n = 97) pasó del 22% al 13%. En el experimento sobre recortes de impuestos, el acuerdo de los conservadores con que los recortes aumentaron la recaudación pasó del 36% al 67% con la corrección (n = 60), y el de los no conservadores, del 31% al 28% (n = 136). En otros dos experimentos, el resultado fue otro: en la segunda versión del experimento sobre armas de destrucción masiva, la corrección hizo a los conservadores más propensos a creer que Irak no tenía esas armas, y los autores dicen que no está claro por qué; en el experimento sobre células madre, la corrección funcionó para conservadores y moderados, y no para liberales, sin backfire. Atribuir la noticia al *New York Times* o a FoxNews.com no alteró el efecto de la corrección en ninguno de los tres experimentos del Estudio 2.

Los autores colocan las salvedades en el propio artículo. Todos los resultados de backfire vinieron de conservadores, y sin más estudios es imposible saber si liberales y conservadores reaccionan de modo distinto a las correcciones; sería valioso replicar con una muestra no universitaria. Un análisis post hoc, que los autores describen como llevar los datos al límite, sugiere que la corrección falló entre los conservadores para quienes Irak era el problema más importante, pero solo 34 encuestados (ocho a la derecha del centro) lo señala-

ron. Y escriben que las percepciones erróneas contemporáneas sobre política son un blanco móvil, que puede cambiar deprisa, como sugiere la diferencia entre los dos experimentos sobre armas de destrucción masiva.

La corrección también llega como advertencia. Freeze et al. (2021) estudiaron lo que llaman efecto de la "verdad contaminada": la advertencia de desinformación aplicada a noticias auténticas, que reduce su credibilidad y perjudica la memoria política. En un experimento en línea con 434 participantes de MTurk, todos vieron un video de cuatro minutos de C-SPAN, con discursos de diputados, respondieron 22 preguntas sin relación con él y fueron distribuidos por sorteo entre seis condiciones. Leían un artículo periodístico fabricado sobre el video, ya vago (control), ya con los hechos correctos, ya con parte de los hechos alterados, y recibían o no la advertencia "parte de la información presentada en el artículo de noticias que usted leyó era inexacta". Los autores llaman válida a la advertencia cuando señala correctamente la desinformación, e inválida cuando recae sobre un artículo sin hechos alterados. Después todos hicieron una prueba de reconocimiento de 20 hechos del video y evaluaron la credibilidad del video y del artículo. La advertencia hizo que el artículo pareciera 0,59 puntos menos creíble (en una escala de 5), en todas las condiciones, incluso cuando el artículo era exacto. El efecto sobre la memoria solo fue significativo en una de las tres pruebas.

DATO	En cuatro experimentos con estudiantes, la corrección incluida en la noticia hizo subir el acuerdo de los conservadores con la afirmación engañosa en dos (armas de destrucción masiva, Estudio 1; recortes de impuestos) y funcionó entre ellos en los otros dos (armas, segunda versión; células madre, para conservadores y moderados y no para liberales). Entre los no conservadores, el acuerdo bajó en los dos temas en que se informan las cifras (Nyhan & Reifler, 2010).
INTERPRETACIÓN	El estudio mide el acuerdo con una afirmación, en una encuesta, justo después de la lectura. No dice qué hace la persona con la afirmación ni cuánto tiempo se mantiene la respuesta. Los resultados de backfire vienen solo de percepciones erróneas sostenidas por conservadores, en los temas elegidos, y los autores dicen no poder afirmar que los campos políticos difieran; el estudio no sostiene ni "los conservadores rechazan las correcciones" ni "las correcciones salen por la culata".
HIPÓTESIS	Los autores leen el backfire como contraargumentación ante información incongruente con la preferencia (interpretación de ellos, sin prueba). Se añaden aquí dos lecturas rivales, igualmente sin prueba en este diseño. La primera es la del Capítulo 6: la corrección es un estímulo aversivo para quien tiene la afirmación ligada a su comunidad, y discrepar más es una respuesta de contracontrol, o de evitación de este. La segunda es la de la repetición (Capítulo 3, sección 3.5): la corrección vuelve a presentar la afirmación, e instrucciones del tipo "no creas que todos los miembros del grupo X son Y" pueden aumentar la saliencia del grupo y añadir atributos a la categoría (Dixon et al., 2003). Separar las tres lecturas exigiría variar si la persona declaró antes su posición en público y si la corrección repite o no el enunciado.

La advertencia retrospectiva válida, dada después de la lectura, ayudó poco. La puntuación media de memoria fue del 59% en el control, el 46% con desinformación y sin advertencia y el 52% con desinformación y advertencia válida, y la puntuación de desinformación pasó del 33% al 25%; la diferencia entre el 46% y el 52% no fue significativa al 0,05 ($p = 0{,}06$). Los autores concluyen que sus hallazgos sugieren evitar las advertencias generales, porque su uso indiscriminado puede reducir la credibilidad de fuentes válidas y llevar a las personas a descartar información útil. Señalan también que la advertencia partió del investigador, sin una fuente claramente especificada, y que sería interesante ver si el razonamiento motivado altera el efecto cuando esta viene de otra fuente.

Leído aquí, el resultado de Freeze et al. tiene la forma de una generalización de estímulos. Una advertencia sin blanco ("esto puede ser desinformación") no dice qué relación del artículo está en juego, y el lector responde reduciendo la confianza en el conjunto, incluso en lo que es válido. Una advertencia específica, que localizara el enunciado y la razón, da ocasión a otra respuesta. Es una hipótesis, y el estudio no la prueba: todas las advertencias fueron generales, y ninguna varió la especificidad. La medida, una vez más, es de juicio: credibilidad y memoria, justo después de la lectura.

Lo que la sección deja como dato es poco, y es útil por eso. Corregir después, en el diseño de los estudios leídos, tiene un efecto incierto, que varió con el tema y no varió con la fuente atribuida a la noticia en los experimentos en que esto se manipuló, y no se midió más allá de la sesión. La hipótesis de que la corrección sea aversiva para quien tiene la afirmación ligada a la comunidad (Capítulo 6) es compatible con los resultados, y ningún estudio la probó. La sección siguiente trata de lo que se hizo para evitar, en lugar de corregir.

12.2 Corregir antes: la inoculación como entrenamiento en discriminación

La inoculación invierte el orden de la corrección. En lugar de desmentir una afirmación después de que circuló, se expone a la persona, antes, a una versión debilitada del ataque. Roozenbeek et al. (2022) definen la inoculación psicológica por dos componentes: una advertencia previa que induce la amenaza percibida de un ataque inminente a las actitudes, y la exposición a una microdosis debilitada de desinformación que trae una refutación preventiva. Los autores atribuyen la teoría a McGuire, en la década de 1960, y esta sigue una analogía con la inmunización médica; no es un concepto del análisis de la conducta, y la lectura que el capítulo hace de ella se propone aquí.

El primer estudio leído es un juego. En *Bad News*, el jugador asume el papel de productor de noticias falsas y aprende seis técnicas de producción de desinformación: polarización, apelación a las emociones, teorías de la conspiración, troleo, desvío de la culpa (el descrédito de opositores, que desvía la atención de acusaciones de sesgo atacando la fuente de la crítica) y suplantación de cuentas. El juego es gratuito, funciona en el navegador y dura unos 15 minutos; el jugador gana seguidores y credibilidad al usar las estrategias y pierde credibilidad o seguidores al elegir opciones alineadas con la conducta periodística ética (Roozenbeek & van der Linden, 2019). La muestra era de conveniencia, reclutada mediante un comunicado de prensa de la universidad,

con 14.266 pares pre-post completos recogidos en tres meses de 2018; era autoseleccionada y sesgada hacia hombres (75%) y hacia personas de algo a muy liberales (59%).

La medida era la confiabilidad, de 1 a 7, de tuits y titulares, evaluada antes y después del juego, en seis ítems, dos de ellos controles de noticias reales. La confiabilidad atribuida bajó en suplantación (de 3,00 a 2,30; d = 0,36), conspiración (de 2,47 a 1,97; d = 0,35) y descrédito (de 2,37 a 1,92; d = 0,30), y los ítems de control prácticamente no cambiaron, lo que los autores leen como indicio de que el juego no vuelve a los participantes simplemente más escépticos. El ítem de polarización, recogido post hoc por una falla técnica de almacenamiento (885 pares), tuvo un efecto menor (d = 0,16). El efecto medio sobre los ítems falsos fue de 2,61 a 2,06 (d = 0,52) y fue mayor entre quienes los juzgaban más confiables en la preprueba.

Los autores señalan las limitaciones. Por razones éticas no aleatorizaron quién jugaría: no hay grupo de control tradicional, y dicen que no pueden descartarse por completo explicaciones alternativas; los estudios futuros deben modelar el decaimiento del efecto, para saber si serían necesarias dosis de refuerzo, y cuándo. Hay una salvedad que añadir: el efecto mayor entre quienes juzgaban los ítems más confiables en la preprueba es el patrón que se espera también de la regresión a la media, y sin grupo de control no hay manera de separarla.

El estudio de 2022 es de otra envergadura. Cinco videos cortos tratan del lenguaje emocionalmente manipulador, la incoherencia, las falsas dicotomías, el chivo expiatorio y los ataques *ad hominem*; cada uno trae advertencia previa, refutación preventiva de la técnica y microdosis en ejemplos inocuos, humorísticos, apolíticos y ficticios. En los estudios 1 a 6, los participantes fueron asignados al azar a ver un video de inoculación de 1,5 minutos o un video neutro, y después evaluaron diez publicaciones sintéticas, manipulativas o neutras, en cuatro medidas, entre ellas la disposición a compartir. El resultado principal es la discriminación (*discernment*): la diferencia entre las medias de los estímulos manipulativos y de los neutros. Los estímulos manipulativos no siempre eran falsos; estaban hechos para usar claramente una técnica de manipulación.

En los estudios 1 a 5 (n = 5.416, muestras por cuotas nacionales de los Estados Unidos), las hipótesis prerregistradas se sostuvieron en 16 de 20 medidas. La d de Cohen para el reconocimiento de la técnica fue de 0,28 (chivo expiatorio) a 0,68 (falsas dicotomías) y, para el compartir, en forma de discriminación, de 0,10 (incoherencia, no significativo) a 0,22 (falsas dicotomías). Un año después, el estudio 6 (n = 1.068) replicó el video de lenguaje emocional, con d =

0,34 para el compartir. Para el efecto menor en los videos de chivo expiatorio e incoherencia, la explicación de los autores es que el grupo de control ya distinguía mejor esas técnicas, es decir, la línea de base era más alta.

En el estudio de campo, dos videos se emitieron como anuncios en YouTube, que los espectadores podían saltar; cerca de 967 mil usuarios lo vieron durante 30 segundos o más, y el 30% de los elegidos al azar recibió, en un plazo de hasta 24 horas, una única pregunta de prueba dentro del propio YouTube, en la que debían identificar la técnica usada en una publicación ficticia (cuatro alternativas, solo una correcta). El grupo de comparación, emparejado por variables demográficas, no vio el video y respondió una de las preguntas; se recogieron 22.632 respuestas. La proporción de aciertos fue mayor en el tratamiento (h = 0,09, por debajo del menor efecto de interés prerregistrado, h = 0,10). Los autores informan un costo medio de cerca de US$ 0,05 por visualización y dicen que los videos se mostraron a cerca de 5,4 millones de personas en 15 días.

Las limitaciones declaradas son las que importan para la lectura. El efecto se midió en un plazo de hasta 24 horas (mediana de 18,4 horas), la duración no se estudió, los videos y las muestras eran de los Estados Unidos, y el estudio de campo evaluó la susceptibilidad, y no la conducta. Cada participante del campo vio un único ítem de prueba, y los pares de estímulos manipulativo y neutro no son espejos exactos, y pueden diferir en otras dimensiones. Los autores resumen, citando un estudio que no se leyó, que inducir la atención a la exactitud llevó a usuarios de Twitter a compartir más fuentes de alta calidad, y dicen que cómo reducir la susceptibilidad en entornos reales sigue siendo una cuestión abierta. No encontraron interacciones consistentes entre la condición y las covariables demográficas e ideológicas.

La lectura propuesta aquí es que la inoculación, en el estudio de 2022, tiene la forma de un entrenamiento en discriminación con ejemplares de la técnica, y no del contenido. La propiedad común a los ejemplos es la técnica (el lenguaje emocional, la falsa dicotomía); el contenido varía y, en la prueba, es otro. Lo que se mide es la diferencia entre la evaluación de una publicación que usa la técnica y la de una publicación neutra, del mismo contenido, y eso es lo que pide el concepto de discriminación. La Figura 12.2 pone el procedimiento en esa forma, con lo que el estudio mide en trazo continuo y lo que se añade aquí en trazo discontinuo.

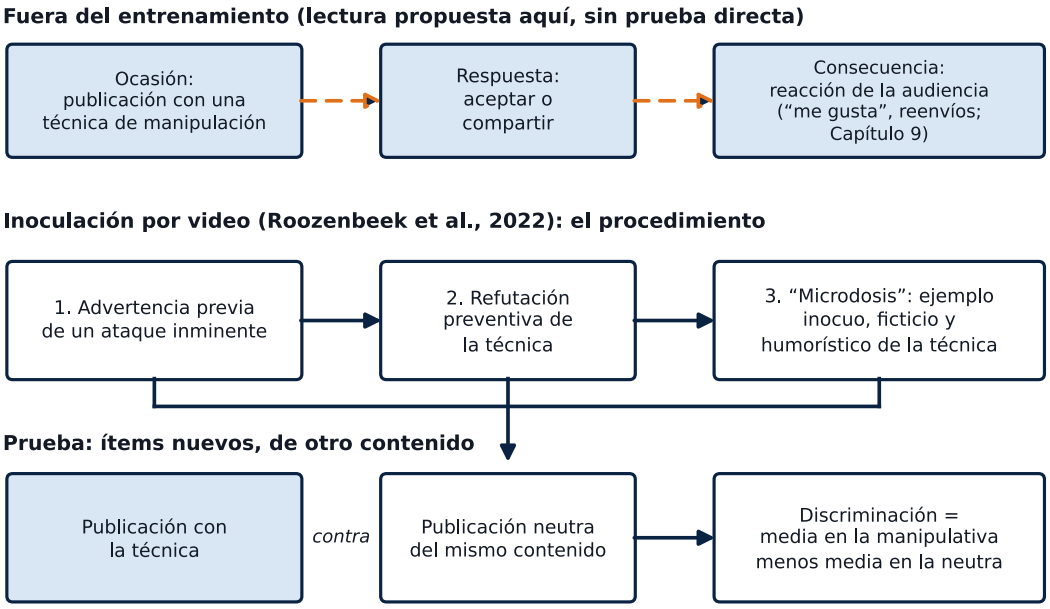
CONCEPTO EN USO

Entrenamiento en discriminación (Skinner, 1957; Shea et al., 2023)

En toda conducta verbal bajo control de estímulos, el estímulo previo establece la ocasión en que la respuesta tiende a ser reforzada y, por un proceso de discriminación operante, pasa a ser la ocasión en que la respuesta tiende a ser emitida (Skinner, 1957). En el tacto, el control se establece reforzando la respuesta de la manera más consistente posible en presencia de un estímulo, con muchos reforzadores diferentes o con un reforzador generalizado, y el control resultante lo ejerce el estímulo. El *entrenamiento con múltiples ejemplares*, en la descripción de Shea et al. (2023), presenta un conjunto de estímulos que muestrea el rango de los estímulos objetivo y de las topografías de respuesta, y continúa hasta que la respuesta quede bajo control de las características de los ejemplos y también de estímulos nuevos que las comparten. Lectura propuesta aquí: el entrenamiento en discriminación es el procedimiento que acerca una respuesta a una propiedad del estímulo, y no al estímulo particular; la prueba de que funcionó es la respuesta a ítems nuevos que comparten la propiedad, con otro contenido.

FIGURA 12.2

Inoculación por video leída como entrenamiento en discriminación



Esquema original, sin datos. Trazo continuo: el procedimiento del estudio. Discontinuo: lectura propuesta aquí, que la medida del estudio no alcanza (el estudio de campo no midió el compartir; la consecuencia social no entra en el entrenamiento).

Nota. Esquema original, sin datos. Trazo continuo: el procedimiento del estudio. Trazo discontinuo: lectura propuesta aquí, que la medida del estudio no alcanza; el estudio de campo no midió el compartir, y la consecuencia social no entra en el entrenamiento.

La lectura tiene dos puntos débiles, que conviene decir. El primero es que, en los videos, no hay consecuencia diferencial programada para discriminar: el espectador ve, oye y responde después. Lo que el procedimiento tiene es instrucción y ejemplo, y el término "entrenamiento" es, en buena parte, una analogía. En *Bad News* hay consecuencia, en seguidores y credibilidad, pero es por usar la técnica, y no por reconocerla; la prueba mide el otro repertorio, el de evaluar, y la transferencia de uno a otro es una suposición. El segundo es que la franja discontinua de la Figura 12.2 es la que importa para el análisis: la técnica aparece, fuera del estudio, en una ocasión en que hay audiencia, y la respuesta de compartir tiene la consecuencia que el Capítulo 9 describió. Nada en el entrenamiento toca esa consecuencia.

DATO	En el juego *Bad News*, sin grupo de control, la confiabilidad atribuida a ítems con tres de las técnicas bajó con d de 0,30 a 0,36, y la de los ítems reales no cambió (Roozenbeek & van der Linden, 2019). En los videos, seis experimentos aleatorizados sostuvieron 16 de 20 hipótesis prerregistradas en los estudios 1 a 5, y el estudio de campo tuvo h = 0,09, por debajo del menor efecto de interés, medido en un plazo de hasta 24 horas, en los Estados Unidos y como susceptibilidad (Roozenbeek et al., 2022).
INTERPRETACIÓN	Una exposición breve a ejemplos de una técnica mejora, justo después, la evaluación de ítems nuevos que la usan, sin volver a la persona simplemente más escéptica. El efecto es pequeño en el campo, y lo que se mide es juicio, y no conducta.
HIPÓTESIS	La inoculación funciona como entrenamiento con múltiples ejemplares de la técnica, y la respuesta de evaluar se generaliza a contenidos nuevos (lectura propuesta aquí). Una lectura rival es que el espectador recibió una regla, y responde a ella mientras la regla es saliente. Las dos predicen el efecto justo después; difieren en la duración y en lo que ocurre cuando la regla compite con la reacción de la audiencia. Separarlas exigiría medir, con seguimiento, el compartir en una plataforma en la que la misma publicación recibe o no reacción.

12.3 Estímulos que señalan precisión

Una tercera familia de intervenciones, además de la corrección (sección 12.1) y de la inoculación (12.2), no enseña antes ni corrige después: pone, en el momento en que la persona lee, un estímulo que señala qué mirar. La que se leyó en detalle es la lista de consejos que Facebook promovió en la parte superior del feed en 14 países, en abril de 2017, desarrollada con la organización First Draft: "Consejos para detectar noticias falsas", con 10 estrategias en la versión

de los Estados Unidos y 6 en la de la India. Guess et al. (2020) la describen como reglas de decisión simples que ayudan a distinguir las noticias principales de las falsas, lo que llaman discriminación, sin exigir que el lector investigue o reflexione a fondo sobre cada noticia.

En los Estados Unidos, el estudio fue un panel en línea de dos olas, realizado por YouGov después de las elecciones legislativas de 2018 (ola 1, N = 4.907; ola 2, N = 4.283), con participantes distribuidos al azar para leer o no los consejos. La medida principal fue la exactitud percibida (de 1 a 4) de titulares principales, hiperpartidistas y falsos, que circularon en fuentes reales o en Facebook o WhatsApp en los seis meses anteriores, parte de ellos clasificados como falsos por al menos una agencia de verificación. En la primera ola, los consejos redujeron la exactitud percibida de las noticias falsas en casi 0,2 puntos en la escala de 4 puntos (ITT = −0,196, el efecto sobre todos los asignados) y en casi 0,3 entre los efectivamente tratados (ATT = −0,299); cerca del 66% superó las preguntas de comprensión.

La intervención tuvo un efecto negativo pequeño sobre la creencia en las noticias principales (ITT = −0,046), pero aumentó la discriminación entre principales y falsas (ITT = 0,146), una mejora del 26,5% respecto del grupo de control. Con una mediana de 20 días entre las olas, el efecto sobre la exactitud percibida de las noticias falsas siguió siendo distinguible de cero, pero se atenuó en más de la mitad (ITT = −0,080). En la medida binaria, la proporción que evaluaba los titulares falsos como muy o algo exactos bajó del 32% en el control al 24% en la primera ola y, en la segunda, del 33% al 29%.

En la muestra en línea de la India, la intervención redujo la exactitud percibida de las noticias falsas (ITT = −0,126) y mejoró la discriminación en un 17,3%; solo el 28% superó las preguntas de comprensión, frente al 66% en los Estados Unidos. En la muestra presencial, aleatoria del padrón electoral de cuatro distritos de Uttar Pradesh y con aplicación oral, los autores no encontraron evidencia de efecto; el 11% de esa muestra usaba WhatsApp, frente al 90% de la muestra en línea. En la segunda ola, no encontraron evidencia confiable de efecto en ninguna de las dos muestras de la India.

Lo que el estudio no encontró importa tanto como lo que encontró. Los autores dicen que no hay evidencia suficiente de que la intervención haya cambiado la intención declarada de compartir noticias falsas ni medidas conductuales de consumo de noticias; las intenciones de compartir noticias principales aumentaron, y las de hiperpartidistas disminuyeron. Entre las salvedades: los efectos fueron modestos y decayeron; hubo un efecto negativo pequeño sobre la exactitud percibida de las noticias principales; muchos usuarios reales de Facebook probablemente ignoraron el enlace; y la ausencia de cambio en el

consumo real tal vez se deba a que los hábitos de información están arraigados. Con datos de monitoreo de navegación, estiman, en un cálculo que describen como hipotético, que los falsos positivos bajarían del 6,1% al 4,9% de las historias consumidas.

Dos comentarios de los autores sitúan la intervención. Comparan el efecto con las respuestas a las etiquetas de "disputado" de Facebook en un estudio anterior (ITT = −0,236) y a una advertencia general (ITT = −0,079); esos estudios no se leyeron. Y señalan que intentar identificar noticias falsas mediante el escrutinio del titular difiere del enfoque de los verificadores profesionales, que usan la lectura lateral de otras fuentes, y que los estudios futuros deben aclarar si el efecto viene de una mayor preocupación por la exactitud o de heurísticas más eficaces.

Hay otros dos resultados de señalización de exactitud en las fuentes leídas. En uno de los estudios de Ceylan et al. (2023), juzgar la exactitud de un titular antes de decidir compartirlo redujo el compartir titulares falsos, pero no aumentó la discriminación de los usuarios más habituales más que la de los menos habituales; el estudio se hizo con 839 participantes, y la tarea era un botón de compartir simulado. Y Freeze et al. (2021), en la sección 12.1, muestran el precio de la señal sin blanco.

DATO	En los Estados Unidos, la lista de consejos redujo en cerca de 0,2 puntos (de 4) la exactitud percibida de las noticias falsas, aumentó en un 26,5% la discriminación, y el efecto bajó en más de la mitad con una mediana de 20 días; en la India, el efecto fue menor en la muestra en línea y no se encontró en la presencial ni en la segunda ola. No hubo evidencia suficiente de cambio en la intención declarada de compartir noticias falsas ni en el consumo observado (Guess et al., 2020).
INTERPRETACIÓN	El consejo cambia el juicio sobre un titular, por poco tiempo y donde se entendió (66% en los Estados Unidos, 28% en la muestra en línea de la India) y donde el medio de circulación es el del estudio (11% de WhatsApp en la muestra presencial, 90% en la en línea). No cambia, en los datos, lo que las personas hacen con las noticias.

HIPÓTESIS El consejo es un estímulo verbal que describe una contingencia: una regla, en el vocabulario del Capítulo 3 (sección 3.7). Las reglas controlan mientras son salientes y la respuesta es barata; sin consecuencia programada por aplicarlas, y con la reacción de la audiencia (Capítulo 9) manteniendo la respuesta de compartir, el decaimiento que los autores encontraron es el esperado, y el hábito, en el sentido de Ceylan et al., deja poco espacio para ellas. Es una lectura propuesta aquí; el mecanismo, dicen los propios autores, está por aclarar.

12.4 Ante quien habla: contradiscurso, compromiso público y autoclíticos de fuente

Entre quien lee y la plataforma que entrega está la interacción: lo que alguien hace, en la misma red, ante quien emitió un enunciado falso o clasificado, por una instancia externa, como discurso de odio. Dos estudios leídos evalúan intervenciones de este tipo. La sección termina con una hipótesis propuesta aquí que ningún estudio prueba.

Hangartner et al. (2021) hicieron un experimento de campo en Twitter. Seleccionaron, mediante un diccionario de términos xenófobos, análisis de sentimiento y clasificación manual, a usuarios que habían enviado tuits xenófobos o racistas en inglés, excluyendo cuentas verificadas, retuits, tuits posiblemente sarcásticos y cuentas probablemente de bots o de menores. Cada usuario fue sorteado para uno de tres contradiscursos de aplicación única (empatía, advertencia de consecuencias, humor) o para el control (40%), sin intervención; el contradiscurso llegaba, en un plazo de hasta 24 horas, desde un perfil neutro controlado por un humano, creado unas cuatro semanas antes, con historial apolítico y cerca de 100 seguidores. El de empatía buscaba despertar empatía (decir, por ejemplo, que es doloroso ver ese tipo de lenguaje); el de advertencia recordaba que amigos y familiares pueden ver el tuit; el de humor usaba un meme de animal con una leyenda que pedía dejar de tuitear. El trabajo de campo fue del 23 de noviembre de 2020 al 17 de enero de 2021, y la muestra analítica, después de la pérdida de 230 cuentas privadas o suspendidas, fue de 1.350 usuarios (humor 284, advertencia 269, empatía 259, control 538).

El análisis fue por intención de tratar, es decir, cada usuario cuenta en el grupo al que fue sorteado, haya recibido o no el mensaje (el 5,5% de los tuits fueron borrados por quienes los enviaron entre el sorteo y la aplicación), con regresiones separadas para cada uno de los seis desenlaces. Como son muchos desenlaces y tratamientos, los autores dan también valores de P ajustados (por el método de Benjamini–Hochberg, que corrige el exceso de resultados

significativos por azar que producen las pruebas múltiples). La empatía redujo el número de tuits xenófobos en las cuatro semanas siguientes en 0,10 desviaciones estándar (P = 0,056; P ajustado = 0,112) y el total de tuits en 0,13 (P = 0,039; ajustado = 0,112), y aumentó la eliminación del tuit original en 0,21 (P = 0,009; ajustado = 0,053), cerca de 8,4 puntos porcentuales. En número de tuits, los usuarios del grupo de empatía enviaron, en promedio, 1,3 tuits xenófobos menos y 91,6 tuits menos en total. La prueba F conjunta, que examina de una vez todos los desenlaces, rechazó la hipótesis nula de ausencia de efecto para la empatía (P = 0,017), pero no para el humor (P = 0,285) ni para la advertencia de consecuencias (P = 0,633).

Los autores dicen que, contrariamente a las hipótesis prerregistradas, no encontraron efectos significativos de la empatía en los otros tres desenlaces, y que el análisis separado de los dos componentes de la proporción de tuits xenófobos es exploratorio; el desenlace prerregistrado, la propia proporción, no cambió. Concluyen que los resultados moderan las expectativas sobre los contradiscursos más comunes usados por las organizaciones y recomiendan estudiar intervenciones repetidas y prestar atención a efectos inhibidores (*chilling*) "más allá del discurso de odio", expresión de ellos. El estudio fue aprobado por un comité de ética y prerregistrado. La premisa, que los autores apoyan en otros trabajos, es que la moderación de contenidos puede suprimir expresiones valiosas e ilícitas y dispersar, en lugar de reducir, lo que llaman discurso de odio, y que el contradiscurso busca reducir el odio mediante la persuasión de quien lo emite, y no suprimir la expresión.

El segundo estudio es de otra naturaleza. Tsipursky et al. (2018) describen el *Pro-Truth Pledge*, una propuesta en la que los firmantes, ciudadanos y figuras públicas, se comprometen con 12 conductas correlacionadas con la orientación a la veracidad, agrupadas en compartir, honrar y alentar la verdad; entre ellas, verificar la información antes de aceptarla y compartirla, retractarse si no puede verificarla y pedir a los aliados que se retracten de la información refutada. En el análisis de los autores, el compromiso funciona como estímulo discriminativo verbal, una regla que describe una contingencia en la que las conductas veraces tendrían consecuencias positivas y las no veraces, negativas. Para las figuras públicas, el diseño prevé la reputación como consecuencia: defensores voluntarios hacen el seguimiento, un comité emite un aviso a la prensa en caso de incumplimiento y la figura aparece en el sitio web como incumplidora; el artículo describe el mecanismo como diseño y no informa de uso ni de sanciones efectivas.

El compromiso se lanzó en diciembre de 2016 y, en mayo de 2018, tenía más de 7.500 firmantes, de los cuales más de 1.200 eran figuras públicas. Los autores reconocen que los resultados de compromisos de este tipo han sido mixtos y

que la falsa señalización, en la que la figura pública firma sin cumplir, puede impedir que el compromiso proporcione información creíble; plantean la hipótesis de que quien firma sin otras formas de participación activa tendrá un impacto a largo plazo pequeño o despreciable. El estudio empírico evaluó conducta pasada, y no la actual, con 21 participantes reclutados mediante publicaciones en Facebook y correos electrónicos a quienes habían firmado y pedido actualizaciones. Dos codificadores calificaron de 1 (la publicación es desinformación) a 5 (orientada a combatir la desinformación y promover la verdad) las diez primeras publicaciones con contenido noticioso hechas cuatro semanas después del compromiso, comparadas con las diez primeras del mismo período del año anterior. La media pasó de 2,49 (DE = 0,60) a 3,65 (DE = 0,41), d = −1,93 (negativo porque el artículo resta la puntuación posterior de la anterior; la evaluación subió), con un alfa de Krippendorff de 0,85, índice de acuerdo entre los dos codificadores.

Los autores escriben que el estudio aporta evidencia de que las personas se comportaron de modo más honesto en Facebook porque firmaron el compromiso, y lo describen como preliminar y voluntario, a replicar con una muestra mayor y grupos de control. Se añade lo que el diseño impone: no hay grupo de comparación; los 21 se autoseleccionaron; la calificación mezcla la calidad de la publicación con la interacción en los comentarios; y el lenguaje causal excede lo que sostiene un pre-post sin control, porque quien decide firmar y pedir actualizaciones es quien está dispuesto a cambiar. Lo que el estudio tiene de aprovechable es la descripción de la contingencia dispuesta, y no el efecto.

Los dos estudios actúan sobre quien habla, después o antes del enunciado, y ninguno trata de lo que el Capítulo 3 identificó como el camino de la declaración falsa en cadena: el intraverbal con forma de tacto. La sección 3.3 describió que la comunidad refuerza la respuesta con la forma de un relato directo del evento, aunque el hablante la emita a partir de lo que oyó; la sección 3.6, que el autoclítico de fuente es lo que diría que la respuesta no es un tacto. Skinner (1957) señala que preguntar "¿Lo viste, o alguien te lo contó?" es preguntar si la respuesta fue un tacto, o una respuesta ecoica o intraverbal a la conducta verbal de otra persona; que en muchas comunidades es de buena educación empezar a hablar con un autoclítico que indique de dónde viene lo que se dice; y que las contingencias de la conducta autodescriptiva las dispone la comunidad cuando tiene razones para preguntar "¿Qué dijiste?", "¿Dijiste eso?" y "¿Por qué dijiste eso?" (pp. 314–316).

La hipótesis propuesta aquí lleva a una intervención que no está en ningún estudio leído: enseñar y reforzar, en la comunidad, la discriminación entre el tacto y el intraverbal con forma de tacto, y la emisión del autoclítico de fuente. Tiene dos brazos. El del oyente consiste en preguntar por la fuente y no re-

forzar, con un "me gusta", un reenvío o el acuerdo, el enunciado que llega sin ella; la pregunta hace que el intraverbal con forma de tacto sea menos reforzado que el enunciado con la fuente marcada. El del hablante consiste en marcar la fuente ("leí en...", "me dijeron que...") y en retractarse, lo que se aproxima a lo que el compromiso de Tsipursky et al. pide al mandar verificar antes de compartir y retractarse. El punto de intervención es el mismo de la sección 12.3, el momento de la lectura, con una diferencia: el estímulo es una consecuencia dispuesta por otra persona, y no una regla impresa. Un diseño en el formato de Hangartner et al. serviría para probarla: un perfil neutro que, ante una afirmación verificada como falsa, pregunta de dónde viene, frente a un control sin respuesta, midiendo la presencia de marcadores de fuente en las publicaciones siguientes, la eliminación del enunciado y el reenvío, en ambos campos políticos.

DATO	Entre 1.350 usuarios de Twitter, el contradiscurso por empatía aumentó la eliminación del tuit en cerca de 8,4 puntos porcentuales y redujo, en 0,10 y 0,13 desviaciones estándar, el número de tuits xenófobos y el total, en cuatro semanas, con valores de P ajustados de 0,053 a 0,112; la proporción de tuits xenófobos, desenlace prerregistrado, no cambió, y el humor y la advertencia no tuvieron un efecto consistente (Hangartner et al., 2021). Entre 21 firmantes autoseleccionados, la calificación de publicaciones en Facebook pasó de 2,49 a 3,65, sin grupo de control (Tsipursky et al., 2018).
INTERPRETACIÓN	Los dos estudios miden conducta observada de quien fue blanco o de quien firmó, con seguimiento de cuatro semanas, y ninguno mide el efecto sobre quien solo lee. En el estudio de campo, los dos componentes de la proporción de tuits xenófobos bajaron en un tamaño similar, en desviaciones estándar, y el resultado es compatible con que el usuario tuitee menos, y no con que tuitee menos odio. El estudio del compromiso es la descripción de una contingencia con un pre-post sin control.
HIPÓTESIS	Un mensaje único de una cuenta desconocida puede funcionar como castigo social leve, y entonces borrar sería escape, y tuitear menos, evitación (lectura propuesta aquí, junto a la persuasión que proponen los autores). Nada en los datos separa las dos. En cuanto al autoclítico de fuente, la hipótesis es la de la sección: el oyente que pregunta por la fuente altera la consecuencia de la forma de tacto. No hay datos que la sostengan ni que la refuten.

12.5 Reducir el prejuicio mediante equivalencia y mediante cómics

El Capítulo 4 mostró que las relaciones evaluativas pueden formarse y reorganizarse en el laboratorio. La pregunta de esta sección es otra: qué dicen las fuentes leídas del uso de esos procedimientos para reducir el prejuicio.

El de Dixon y Lemke (2007) es el estudio que la literatura leída cita como modelo, y vale la pena leerlo por lo que hizo. Participaron 12 estudiantes de grado (blancos o afroamericanos) de la Southern Illinois University, por crédito extra; los autores definen operacionalmente el prejuicio como responder de manera diferencial a los miembros de una raza específica. En la preprueba, cada participante evaluó 30 imágenes (diez de hombres blancos estadounidenses, diez de hombres de Oriente Medio y diez de objetos comunes), de 1 (malvado) a 10 (bueno). El entrenamiento fue de igualación a la muestra, con consecuencia diferencial (sonido agradable para el acierto, sonido aversivo para el error) y pruebas de simetría y equivalencia en extinción; la relación entrenada unía la imagen del hombre estadounidense blanco a la palabra “malvado”, la del hombre de Oriente Medio a la palabra “bueno” y la del objeto común a la palabra “neutral” (p. 10). Cuatro participantes evaluaron todos los estímulos masculinos con “5” en la preprueba y sus datos se excluyeron; los ocho restantes evaluaron a los hombres de Oriente Medio más cerca de “malvado” que a los estadounidenses.

Todos completaron el entrenamiento, y siete de ocho alcanzaron el 90% o más en cada prueba de simetría y equivalencia. Los autores escriben que el hombre de Oriente Medio pasó a ser evaluado más cerca de “bueno” y el estadounidense más cerca de “malvado”, y que el patrón se generalizó a nueve estímulos nuevos. La Tabla 1 del artículo permite un recuento que los autores no hacen: en cinco de los ocho participantes (1, 3, 5, 8 y 12), la media para el hombre de Oriente Medio quedó por encima de la del hombre estadounidense en la posprueba, y la media para el estadounidense bajó a valores de 1,00 a 1,50. En la discusión, dicen que no todos los que alcanzaron el criterio de equivalencia mostraron diferencias significativas y atribuyen parte de la ausencia de cambio a contingencias sociales culturales implícitas, fuente de control mayor que la diferenciación por puntos usada en el estudio; los objetos comunes, previstos como neutros, recibieron las calificaciones más altas en la preprueba y en la posprueba. Advierten que los participantes pueden no responder conforme a su red relacional verdadera, por contingencias sociales que redu-

cen la exhibición de conducta prejuiciosa, y registran que, cuando se demuestra equivalencia entre estímulos con relaciones preexistentes fuertes, puede ser difícil alterar esas relaciones.

Dos puntos de lectura. El primero es lo que el entrenamiento enseñó. Según la definición del Capítulo 4, una actitud en la que los atributos evaluativos son negativos es un prejuicio; el procedimiento de Dixon y Lemke relacionó "malvado" con el hombre estadounidense blanco y "bueno" con el de Oriente Medio, de modo que el resultado esperado no es la neutralización, sino la inversión del signo, y la inversión instala en el laboratorio una evaluación negativa de otro grupo. Es una lectura propuesta aquí, y vale como cuidado de diseño, y no como crítica a los autores, que lo describen como demostración. El segundo es el recuento: cinco de ocho es una mayoría sin unanimidad, en una muestra pequeña, sin grupo de control y sin seguimiento. En niños, con rostros de personas negras y un símbolo positivo, y sin clase negativa para el otro grupo, el Capítulo 4 relata que 13 de 13 formaron las clases y nueve mantuvieron las relaciones (Mizael et al., 2016a), y que, en 46 niños, 33 formaron clases y 27 mantuvieron al menos una relación después de seis semanas, con reducción también en el grupo control, en menor grado (Mizael et al., 2021b). Mizael et al. (2016b) sostienen que un entrenamiento de una o dos horas es probablemente insuficiente para modificar relaciones preexperimentales ampliamente reforzadas por la comunidad verbal; es una hipótesis de los revisores.

Los cómics entran aquí por ser el procedimiento que usan los estudios del Instituto Walden4, y por una razón que interesa a la intervención: los autores de uno de los volúmenes argumentan que las narrativas, al tratar el prejuicio de forma sutil, diluida en la trama, minimizan la percepción de intención persuasiva y permiten que el lector se identifique con los personajes, lo que los métodos tradicionales de cambio de actitud no logran ante la resistencia actitudinal (Moreira et al., 2022a). El procedimiento, el elenco y las escenas están en las Figuras 4.15 a 4.18 del Capítulo 4. El volumen editado por Moreira (2025a) reúne estudios de sujeto único, mayoritariamente con universitarios, con igualación a la muestra y con cómics, sobre sexismo, homofobia, edadismo, transfobia y gordofobia, entre otros, en los que la actitud se mide con una escala de diferencial semántico; el resumen relata que la igualación formó clases y transfirió funciones evaluativas, en especial a estímulos inicialmente neutros o con menor carga emocional, y que hubo resistencia al cambio de prejuicios profundamente arraigados o asociados a una fuerte carga emocional. El Experimento 8 de Moreira et al. (2022c) hizo que cinco adultos leyeran una historia con tres personajes, de ideales políticos de izquierda, de derecha y neutro, y evaluaran a los personajes en tres momentos; hubo alteración en las escalas de los cinco, en los tres personajes, en un diseño en el que el sujeto

es su propio control, en preprueba y posprueba. A ese estudio se aplica el criterio aplicado a los otros: no hay grupo de control, no hay seguimiento, la medida es una escala respondida justo después y el objeto de la evaluación es un personaje ficticio, de modo que la alteración de las marcas no se distingue, por sí sola, de un efecto de repetición de la medida (Capítulo 4, sección 4.7).

Tres revisiones sitúan el conjunto. Matsuda et al. (2020) buscaron, en 14 revistas analítico-conductuales, palabras clave de racismo y prejuicio en el título (PsycInfo, desde 1990), 30 años de literatura, y relatan que el resultado fue cero; ampliando la búsqueda a artículos conceptuales y a estudios funcional-contextuales, escriben que hay un déficit notable del análisis de la conducta sobre el racismo y poca investigación para orientar el trabajo. Resumen que los estudios de equivalencia muestran que el sesgo existe y puede medirse, y que los sesgos establecidos son difíciles de superar, aunque hay procedimientos prometedores; advierten que no afirman que tales procedimientos cambien los prejuicios a largo plazo ni que prevengan la conducta racista explícita fuera del laboratorio. Citan, como relatos que no se leyeron en el original, estudios de atención plena y de toma de perspectiva, un piloto de clase basada en ACT y un estudio con docentes, en el que, después de automonitoreo, retroalimentación de desempeño y entrenamiento, niños de todas las categorías recibieron tasas equitativas de elogio y reprimenda. Concluyen que no presentan una solución definitiva, porque la investigación conductual actual no es suficiente ni robusta, y recomiendan extender y replicar estudios, estudiar resultados longitudinales y desarrollar intervenciones comunitarias.

Shea et al. (2023) revisan, en un artículo narrativo sin método de búsqueda descrito, las estrategias comunes contra el prejuicio (educación, entrenamiento en diversidad y vergüenza) desde la Teoría de los Marcos Relacionales, porque sus resultados son limitados. Relatan, a partir de trabajos que no se leyeron en el original, cambios pequeños y autoinformados en las actitudes, que decayeron con el tiempo, en un metaanálisis de 230 muestras en 40 años; un entrenamiento de prevención del acoso con un pequeño efecto negativo en las actitudes de 90 universitarios varones; una literatura de entrenamiento en diversidad fragmentada y limitada por medidas pobres y falta de datos de mantenimiento y de conducta; y un procedimiento de vergüenza sin efecto sobre atribuciones relativas a discapacidades. La hipótesis de los autores es que el modo de impartir la educación puede fortalecer el prejuicio al volver a presentar las visiones prejuiciosas, y que la vergüenza puede funcionar como operación motivadora condicionada reflexiva que lleva a esconder el prejuicio. Concluyen que las investigaciones en Teoría de los Marcos Relacionales, en ACT y en IRAP han mostrado resultados prometedores, pero que es imprescindible la investigación longitudinal, y que los tratamientos no deben perpe-

tuar las redes relacionales prejuiciosas; recomiendan, como hipótesis, el entrenamiento con múltiples ejemplares con grupos marginados antes de que los niños sean expuestos a redes que promueven el prejuicio.

Levy et al. (2022) añaden una restricción de método. Dicen que la literatura analítico-conductual muchas veces deja de informar variables demográficas suficientes, incluidas las de raza y las de poblaciones y contextos culturalmente diversos, lo que puede llevar a una comprensión escasa de la eficacia de medidas y evaluaciones. Los estudios de esta sección, en general, informan poco sobre quiénes son los participantes y para quién vale el resultado.

DATO	En el laboratorio, el entrenamiento en equivalencia alteró evaluaciones en el sentido entrenado en cinco de ocho adultos (recuento propio sobre Dixon y Lemke, 2007), en 13 de 13 niños con mantenimiento en nueve (Mizael et al., 2016a) y en 33 de 46 con mantenimiento parcial a las seis semanas (Mizael et al., 2021b). Una búsqueda en 14 revistas no encontró estudios con los términos racismo o prejuicio en el título (Matsuda et al., 2020), y las revisiones dicen que no hay datos de efecto a largo plazo ni de conducta explícita fuera del laboratorio (Matsuda et al., 2020; Shea et al., 2023).
INTERPRETACIÓN	Las relaciones evaluativas son entrenables y, en parte, se mantienen durante semanas, en muestras pequeñas y con medida de escala. Nada de lo leído muestra cambios en lo que las personas dicen o hacen en otros contextos, ni en lo que la comunidad refuerza.
HIPÓTESIS	Las relaciones antiguas siguen reforzadas por la comunidad verbal fuera del procedimiento, y por eso el entrenamiento puntual tiende a ser insuficiente, como sostienen Mizael et al. (2016b) y Shea et al. (2023) (hipótesis de los autores). De ello se extraen dos lecturas: que la intervención que cambia solo la relación en el individuo, sin tocar lo que la comunidad refuerza, es la que menos sostienen las fuentes; y que la forma de la entrega (corregir, instruir, avergonzar) puede volver a presentar la relación que se quiere deshacer, como la sección 12.1 planteó para la corrección.

12.6 Cambiar las contingencias de la plataforma

Esta sección trata del nivel de la plataforma. Lazer et al. (2018) discuten dos estrategias generales de mitigación: capacitar a los individuos para identificar noticias falsas e intervenir en las plataformas para reducir la atención que reciben; y señalan que la cooperación de las plataformas con los investigadores

es crucial para dimensionar el problema y la eficacia de las intervenciones. Las secciones anteriores trataron de la primera. Esta trata de la segunda, que actúa sobre lo que el Capítulo 8 describió como regla de entrega y el Capítulo 11 llamó alcance. Es también la que más depende del acceso: los estudios leídos son pocos, y la mayor parte se hizo con la cooperación de la propia plataforma.

Lo que hace la regla de entrega lo midieron Huszár et al. (2022), en Twitter. Con la personalización de la línea de tiempo, introducida en 2016, todos los partidos tuvieron amplificación positiva (la razón entre el alcance en el grupo con la línea de tiempo personalizada y en el grupo de control, que siguió con la cronológica; 0% es alcance igual), en algunos casos por encima del 200%, es decir, una audiencia tres veces la de la línea de tiempo cronológica; con excepción de Alemania, el mayor partido de derecha o centroderecha tuvo una amplificación mayor que el de izquierda o centroizquierda, y, a nivel individual, una prueba de permutación no detectó una asociación estadísticamente significativa entre el partido del político y su amplificación. No es una intervención evaluada, sino la medida de una regla que ya estaba en funcionamiento. El grupo de control, del 1% de los usuarios, no se creó para la investigación, y los propios autores dicen que no está aislado de efectos indirectos, de modo que las estadísticas probablemente subestiman los efectos; el trabajo fue revisado por los equipos jurídico y de privacidad de Twitter, y no por un comité de ética.

Nyhan et al. (2023) hicieron lo que Huszár et al. no pudieron: una intervención aleatorizada en el feed. Los 23.377 participantes, adultos de los Estados Unidos, fueron reclutados mediante invitaciones en la parte superior del feed de Facebook, dieron su consentimiento y respondieron encuestas en olas antes, durante y después de la elección. Para los asignados al tratamiento, todo el contenido de amigos, grupos y páginas que se preveía que compartían la inclinación política del participante fue relegado en el feed, del 24 de septiembre al 23 de diciembre de 2020, con la mayor intensidad de relegación que una prueba previa indicó que no dejaría el feed casi vacío; el feed del grupo de control no se alteró sistemáticamente. La exposición a fuentes de la misma inclinación bajó del 53,7% al 36,2%; la exposición a fuentes transversales subió del 20,7% al 27,9%, y la de fuentes ni de la misma inclinación ni transversales, del 25,6% al 35,9%. No hubo efecto medible en ocho medidas de actitud preregistradas, como la polarización afectiva, el extremismo ideológico, la evaluación de candidatos y la creencia en afirmaciones falsas, y los autores descartan efectos de ±0,12 desviaciones estándar o más, es decir, de poco más de una décima parte de la variación típica de cada medida. Quienes tuvieron la exposición reducida interactuaron más, por exposición, con el contenido de la

misma inclinación que aún veían. Entre las explicaciones posibles del resultado nulo, los autores citan que el contenido político afín es una fracción de lo que se ve, que Facebook es una fracción de la dieta informativa, que persuadir es difícil y que el diseño no capta efectos del uso previo ni efectos acumulados a lo largo de años; observan además que reducir la exposición podría inducir a las personas a buscar esa información en otro lugar de internet.

El tercer estudio trata de la moderación. Álvarez-Benjumea y Winter (2018) pidieron a 180 participantes, reclutados en una plataforma de crowdsourcing, que comentaran, en un foro simulado, nueve fotos de temas controvertidos (como el feminismo, los derechos LGBT y los refugiados), en cuatro condiciones: base (dos comentarios amistosos, dos neutros y dos hostiles), censurada (contenido de odio borrado), extremadamente censurada (solo comentarios amistosos) y contradiscurso (comentarios hostiles con réplicas que destacan su inaceptabilidad). Los comentarios hostiles (calificación 7 o más) fueron 24 (5,88%) en la base, 4 (0,99%) en la condición censurada, 18 (4,44%) en la de contradiscurso y 15 (4,10%) en la extremadamente censurada; solo la reducción de la condición censurada fue significativa. Los autores sugieren que la conducta normativa en línea está motivada por normas descriptivas, lo que hacen los demás, más que por normas prescriptivas, y que las réplicas, al ser ellas mismas hostiles, pueden añadir ambigüedad sobre la norma. Advierten que no tienen datos sobre la disyuntiva entre censura y libertad de expresión y que el anonimato del foro impide procesos de identificación grupal.

Ceylan et al. (2023), que el Capítulo 9 leyó, entran aquí por dos motivos. El primero es el estudio 4, que funciona como modelo experimental de una regla de entrega: en 80 ensayos, los participantes ganaban “+5 puntos” cuando la elección seguía el criterio del entrenamiento (en el de exactitud, compartir un titular verdadero o no compartir uno falso; en el de desinformación, lo inverso) o no recibían ninguna retroalimentación (control); en la prueba, sin retroalimentación, el patrón persistió. Después del entrenamiento en exactitud, se compartió el 66% de los verdaderos y el 24% de los falsos; en el control, el 40% y el 24%; después del entrenamiento en desinformación, el 54% y el 33%. Lo que hizo la retroalimentación fue, en el primer caso, aumentar el compartir lo verdadero sin reducir el compartir lo falso, y, en el segundo, aumentar el compartir en general. El segundo motivo es la conclusión de los autores: los usuarios habituales son parte integral de los modelos de lucro de las redes basados en anuncios y, por eso, difícilmente las plataformas crearán estructuras de recompensa que impidan los hábitos.

DATO	La personalización de la línea de tiempo amplificó a todos los partidos, con más amplificación a la derecha que a la izquierda en los mayores partidos de seis de los siete países y sin asociación detectable a nivel del político (Huszár et al., 2022). Relegar las fuentes de la misma inclinación en Facebook durante tres meses redujo esa exposición del 53,7% al 36,2% y no tuvo un efecto medible en ocho medidas de actitud (Nyhan et al., 2023). Borrar los comentarios hostiles en un foro simulado redujo los comentarios hostiles del 5,88% al 0,99% (Álvarez-Benjumea & Winter, 2018).
INTERPRETACIÓN	La regla de entrega cambia lo que las personas ven, y la moderación cambia lo que se publica en un foro simulado; pero, en el único experimento aleatorizado del feed leído, una reducción de 17,5 puntos porcentuales en la exposición a fuentes de la misma inclinación no cambió las actitudes medidas. Exposición y actitud son variables distintas (Capítulo 11).
HIPÓTESIS	Lo que la regla de entrega altera es la ocasión (el alcance) y, por medio de la reacción de la audiencia, la consecuencia de quien produce (Capítulo 9). Si es así, la intervención que alcanza lo que mantiene la producción es la que cambia el retorno que recibe el productor, y ningún estudio leído la probó: todos midieron lo que la audiencia ve o hace, y ninguno, lo que quien publica pasó a publicar. Probarla exigiría variar el retorno (alcance o reacción visible) y medir el contenido publicado después, con el mismo criterio para productores de cualquier campo político.

La Tabla 12.1 reúne las once intervenciones de las secciones 12.1 a 12.6, con el diseño, la medida, el resultado y el límite que declara el propio estudio.

TABLA 12.1

Las intervenciones leídas: diseño, medida, resultado y límite

Estudio	Intervención y diseño	Medida	Resultado	Límite
Nyhan y Reifler (2010)	Corrección incluida en una noticia simulada; cuatro experimentos entre sujetos, estudiantes	Acuerdo con la afirmación, justo después de la lectura	Backfire en dos de cuatro, entre conservadores; la corrección funcionó en los otros dos	Backfire solo en conservadores; sin seguimiento; muestra de estudiantes

Estudio	Intervención y diseño	Medida	Resultado	Límite
Freeze et al. (2021)	Advertencia de desinformación, inválida o válida; experimento en línea, 434 participantes	Credibilidad y memoria	La advertencia redujo la credibilidad en 0,59 puntos (de 5), incluso la de un artículo exacto; advertencia válida, del 46% al 52% de memoria (p = 0,06)	Advertencia general, del investigador; solo la sesión
Roozenbeek y van der Linden (2019)	Juego Bad News; pre-post sin control, 14.266 pares	Confiabilidad de 1 a 7	d de 0,30 a 0,36 en tres técnicas; ítems reales sin cambios	Sin control; autoseleccionada, 75% hombres; sin seguimiento
Roozenbeek et al. (2022)	Videos de inoculación; seis experimentos aleatorizados y uno de campo (22.632)	Reconocimiento de la técnica, confiabilidad, disposición a compartir	Hipótesis sostenidas en 16 de 20 medidas; campo, h = 0,09	Hasta 24 horas; Estados Unidos; susceptibilidad, no conducta
Guess et al. (2020)	Lista de consejos de Facebook; panel de dos olas (EE. UU. e India) y muestra presencial	Exactitud percibida, intención de compartir, consumo observado	EE. UU.: discriminación +26,5%; el efecto cae en más de la mitad en 20 días; India presencial, sin efecto	Sin evidencia en intención ni en consumo
Dixon y Lemke (2007)	Entrenamiento en equivalencia con valencia invertida; 12 estudiantes, 8 analizados	Evaluación de 1 a 10 de imágenes	Cinco de ocho invirtieron el orden (recuento propio)	Sin control ni seguimiento; objetos con las calificaciones más altas
Mizael et al. (2021b)	Entrenamiento con parámetros diferentes; 46 niños; seguimiento de seis semanas	Formación y mantenimiento de clases; sesgo	33 de 46 formaron clases; 27 mantuvieron al menos una relación	Asociación, no causa; el control también redujo el sesgo
Hangartner et al. (2021)	Contradiscurso de aplicación única por un perfil neutro; 1.350 usuarios de Twitter; cuatro semanas	Tuits xenófobos creados y eliminados	Empatía: eliminación +0,21 DE (ajustado, 0,053); tuits xenófobos −0,10 DE y total −0,13 DE (ajustado, 0,112); desenlace prerregistrado sin efecto	No mide a quien solo lee; humor y advertencia sin efecto
Tsipursky et al. (2018)	Compromiso público (Pro-Truth Pledge); pre-post, 21 firmantes	Calificación de dos codificadores, de 1 a 5, de diez publicaciones	De 2,49 a 3,65 (d = −1,93)	Sin control; autoselección; conducta pasada
Álvarez-Benjumea y Winter (2018)	Eliminación de comentarios hostiles; foro simulado; 180 participantes	Comentarios hostiles (calificación 7 o más)	5,88% en la base y 0,99% en la censurada; contradiscurso sin efecto	Sin datos sobre la disyuntiva con la libertad de expresión

Estudio	Intervención y diseño	Medida	Resultado	Límite
Nyhan et al. (2023)	Relegar fuentes de la misma inclinación en el feed de Facebook; tres meses; n = 23.377	Exposición y ocho medidas de actitud	Exposición del 53,7% al 36,2%; sin efecto en las actitudes (±0,12 DE descartado)	Efectos acumulados no captados; Estados Unidos, 2020

Nota. Elaboración propia, a partir de los artículos; las cifras y las páginas están en las secciones 12.1 a 12.6. La columna "Límite" reúne salvedades de los autores y, en algunos casos, lo que el diseño impone y el texto discute; los recuentos propios vienen señalados.

12.7 Diseño cultural y los límites éticos de quien diseña

De los cuatro niveles de intervención del capítulo, el de la cultura es el único en el que ningún estudio leído evaluó una intervención, la fila vacía de la Figura 12.1. No es por falta de concepto. Glenn (2004) trata, en las páginas finales de su artículo, de lo que se puede diseñar cuando el objeto es una práctica y no una persona, y es suyo el concepto que el análisis necesita.

La lectura permite reordenar lo que el capítulo vio. Las intervenciones de las secciones 12.1 a 12.3 y 12.5 actúan sobre la ocasión de cada individuo, mostrando, enseñando o advirtiendo, y suponen, en el vocabulario de Glenn, que cambiar personalmente la conducta de muchos cambiaría el efecto acumulativo; las de la sección 12.4 actúan sobre la consecuencia que una persona dispone para otra; las de la 12.6 actúan sobre el ambiente que selecciona lo que circula. La hipótesis de los Capítulos 8 y 9 es que lo que mantiene el habla está en la consecuencia, y no en la ocasión. Si es así, la evidencia está donde es barato medir, y las hipótesis están donde está lo que mantiene. Es la observación central del capítulo, y es interpretación: nada en los datos muestra que la intervención sobre la consecuencia funcione mejor, porque ningún estudio leído la probó.

Levy et al. (2022) dan al diseño una condición que las otras fuentes no dan. Describen la humildad cultural como un valor que guía la interacción con personas de otros contextos culturales, un patrón continuo de conducta, y no un resultado terminal alcanzable, y critican la "competencia cultural" por sugerir que el aprendizaje tiene un punto final. Escriben que las conductas que componen el racismo se enseñan, se imitan y se mantienen, de modo incidental o por contingencias dispuestas deliberadamente, y que el racismo existe en los niveles individual, sistémico y cultural. Mencionan las metacontingencias seleccionadas para mantener sistemas racistas y dicen que examinarlas queda fuera del alcance del artículo. Y dicen que el análisis de la conducta, con raíces eurocéntricas y una falta observada de diversidad, equidad e inclusión, no está exento de evaluación, y que necesita adaptarse a la variabilidad cultural

de sus miembros y de las personas atendidas. Lectura propuesta aquí: quien diseña es parte de la práctica que diseña, y la ciencia que ofrece las herramientas también.

CONCEPTO EN USO

Diseño cultural (Glenn, 2004)

Glenn escribe que la ingeniería cultural aún no ha encontrado una base científica segura, que el primer paso es entender los fenómenos que se van a diseñar, y que puede incidir sobre dos tipos de fenómeno: la macroconducta y las metacontingencias. La única manera de actuar sobre los efectos acumulativos de la macroconducta es alterar la conducta del mayor número posible de participantes individuales; cuando el número es grande, el cambio en un pequeño porcentaje puede marcar una diferencia importante. Para diseñar intervenciones, es preciso identificar las contingencias operantes que explican la conducta de los individuos que participan en la práctica, lo que implica modificar las contingencias operantes que mantienen la conducta de muchas personas. Las contingencias conductuales entrelazadas recurrentes pueden cambiarse de dos maneras: alterando el ambiente selector externo y esperando variaciones, o alterando sus componentes para que se adapten mejor al ambiente selector actual. Para Glenn, sin entender cómo surgen y evolucionan las culturas, será difícil tomar decisiones sabias sobre lo que puede cambiarse o debe cambiarse. Lectura propuesta aquí: el diseño cultural es el conjunto de las intervenciones que cambian contingencias operantes de muchos participantes o el ambiente que selecciona las prácticas, y no la información que recibe cada uno.

De ahí los límites éticos, que el capítulo reúne en cinco, todos extraídos de las fuentes y del criterio adoptado aquí. El primero es quién decide qué es falso o qué se clasifica como discurso de odio. El Capítulo 2 mostró que la etiqueta de falso depende de quién verifica, con qué evidencia y en qué momento, y el Capítulo 5, que "discurso de odio" entra en este libro solo como etiqueta externa; diseñar la reducción de algo que define un tercero hace del tercero parte del diseño. El segundo es la libertad de expresión. Zielinski (2021) advierte que las acciones correctivas necesitan salvaguardas de derechos humanos para no llevar a la censura y a la restricción de la libertad de expresión, y que la tecnología sola no basta; Hangartner et al. (2021) parten de la premisa de que la moderación puede suprimir expresiones valiosas y dispersar el odio, y Álvarez-Benjumea y Winter (2018) dicen no tener datos sobre la disyuntiva entre censura y libertad de expresión.

El tercer límite es el del efecto que no es el buscado. La empatía de Hangartner et al. redujo también el total de tuits, y los autores piden atención al efecto inhibidor más allá del discurso de odio; la advertencia general de Freeze et al. redujo la credibilidad de noticias exactas; el entrenamiento en equivalen-

cia de Dixon y Lemke invirtió la valencia; la corrección de Nyhan y Reifler hizo subir, entre los conservadores, el acuerdo con la afirmación en dos de los cuatro experimentos. Toda intervención tiene que leerse también por lo que instala junto a lo que elimina. El cuarto es el del consentimiento y el control por terceros. El trabajo de Huszár et al. fue revisado por los equipos jurídico y de privacidad de la empresa, en el curso de sus operaciones ordinarias y no por un comité de ética, y una evaluación de impacto en la protección de datos concluyó que no eran necesarios mecanismos adicionales de aviso y consentimiento. El de Hangartner et al. fue aprobado por un comité de ética y prerregistrado, y la intervención vino de perfiles creados para el estudio, con un historial neutro y apolítico; el texto leído no menciona el consentimiento de los usuarios. Aquí no se decide entre los dos modos; se registra que, en los dos, la decisión sobre aviso y consentimiento correspondió a quien condujo o revisó el estudio.

El quinto es el del criterio simétrico. Lo que el Capítulo 10 estableció para la ventana de Overton vale aquí: el criterio es el mismo para lo que ensancha y para lo que estrecha, para el campo político de quien emite, para el idioma y para la plataforma, y la salvedad de que cada estudio mide lo que mide es la misma. Una intervención cuyos efectos solo se miden para el contenido de un campo no supera el criterio, cualquiera que sea el campo.

Para quien quiera usar el capítulo en la lectura de una intervención, de una noticia sobre ella o de un informe de plataforma, las preguntas que sugieren las secciones son las siguientes.

1. ¿En qué nivel actúa (individuo, interacción, plataforma o cultura) y sobre qué término de la contingencia (ocasión, respuesta o consecuencia)?
2. ¿Qué se midió: juicio, intención declarada o conducta observada? ¿Con qué seguimiento, y en quién (quien lee, quien habla o quien publica)?
3. ¿Cuál era la comparación, y la asignación fue al azar? ¿Qué dicen los autores que el diseño no permite afirmar?
4. ¿El resultado vale para contenido de otro campo político, de otro idioma y de otra plataforma? ¿El criterio fue el mismo para todos?
5. ¿Qué instala o enseña la intervención además del blanco: una señal sin blanco, una valencia invertida, una regla que compite con la reacción de la audiencia?
6. ¿Quién decide qué cuenta como falso o como discurso de odio, y qué salvaguarda hay contra la restricción de la expresión?

Síntesis

El capítulo organizó las intervenciones en cuatro niveles y leyó once estudios por lo que cada uno midió. Ocho midieron juicio o desempeño en una tarea, cinco midieron conducta observada, cuatro, conducta con seguimiento de una semana o más, y ninguno evaluó una intervención a nivel cultural (Figura 12.1). La corrección después de la lectura tiene un efecto incierto y depende del tema; el backfire de Nyhan y Reifler vino de percepciones erróneas de conservadores, en los temas elegidos, y la advertencia general redujo la credibilidad de noticias exactas (12.1).

La inoculación, leída como entrenamiento en discriminación con ejemplares de la técnica, mejora, justo después, la evaluación de ítems nuevos que la usan, con un efecto pequeño en el campo y sin medida de conducta (12.2). Los consejos para identificar noticias falsas cambiaron el juicio por poco tiempo, donde se entendieron, y no cambiaron, en los datos, el consumo ni, con evidencia suficiente, la intención declarada de compartir (12.3). Ante quien habla, el contradiscurso por empatía aumentó la eliminación y redujo la actividad, sin efecto en el desenlace prerregistrado, y el compromiso público tiene un pre-post sin control (12.4); la hipótesis propuesta aquí, la de enseñar y reforzar el autoclítico de fuente, no tiene prueba.

El entrenamiento en equivalencia altera evaluaciones en el laboratorio, en parte y durante semanas, y ninguna fuente leída muestra cambios en lo que se dice o hace fuera de él; las revisiones piden seguimiento e intervención comunitaria, y una de ellas advierte que volver a presentar la relación puede fortalecerla (12.5). La regla de entrega cambia lo que se ve, y relegar las fuentes de la misma inclinación no cambió las actitudes medidas; la moderación redujo los comentarios hostiles en un foro simulado, y ningún estudio midió lo que quien publica pasó a publicar (12.6). Para Glenn, actuar sobre efectos acumulativos exige cambiar las contingencias operantes de muchos participantes o el ambiente que selecciona las prácticas; los límites éticos son el de quién define qué es falso, la libertad de expresión, el efecto colateral, el consentimiento y la simetría del criterio (12.7).

Lo que el capítulo deja es una asimetría. La evidencia está donde es barato medir, en el juicio de una persona ante una pantalla, y las hipótesis están donde, según el análisis de los Capítulos 8 y 9, está lo que mantiene el habla: en la consecuencia que la audiencia y la plataforma disponen para quien la produce. La pregunta que queda para quien investiga es la de un diseño que mida la conducta de quien publica, ante un cambio de retorno, con seguimiento y con el mismo criterio para todos los campos.

Referencias

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), Artículo 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), Artículo 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. 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), Artículo 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), Artículo 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. (Ed.). (2025a). *Formação e mudança de atitudes: Pareamento ao modelo e história em quadrinhos* (1.ª 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* (1.ª 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* (1.ª 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), Artículo 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), Artículo 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

APÉNDICE

A

Ficha de análisis del discurso

La ficha es la del Capítulo 3 (sección 3.2), lista para usar. Existe para obligar a separar lo que se sabe de lo que se infiere: los cinco primeros campos son dato, los tres siguientes son interpretación y los dos últimos son hipótesis. Donde la fuente no documenta, se escribe "no documentado"; el campo no se llena por inferencia. La misma ficha, con el mismo criterio, vale para declaraciones de cualquier campo político.

#	Campo	Qué registrar	Naturaleza
1	Declaración	El enunciado, transcrito como lo trae la fuente, entre comillas, con la referencia de la fuente	Dato
2	Fecha	Fecha de la declaración	Dato
3	Contexto y medio	Dónde y en qué medio se dijo; la pregunta, el argumento o el evento inmediatamente anterior, si la fuente los trae	Dato
4	Audiencia presente	Quién estaba presente y cómo reaccionó, cuando la fuente lo documenta	Dato
5	Verificación	Agencia, etiqueta, fecha y justificación; la falsedad es siempre de la agencia que la clasificó (Capítulo 2)	Dato
6	Operante probable	Mando, tacto, intraverbal, tacto distorsionado, mando disfrazado o tacto impuro, función ritual o de identidad; y por qué (Tablas 3.1 y 3.4)	Interpretación
7	Variables de control candidatas	Ocasión, audiencia, consecuencia probable, historia de reforzamiento, lo que el oyente haría después	Interpretación
8	Marcas de forma	Autoclíticos de fuente y de certeza, cuantificadores, repetición (pura o con variación)	Interpretación

#	Campo	Qué registrar	Naturaleza
9	Qué cambiaría la declaración	La observación que decidiría entre las lecturas: la ocasión, la audiencia, la comparación entre audiencias o antes y después de la corrección	Hipótesis
10	Interpretaciones rivales	Las seis de la Tabla 3.4, con lo que separaría cada una: control débil por el evento, intraverbal adquirido de otros, tacto distorsionado, mando disfrazado o tacto impuro, función ritual o de identidad, efecto de la selección por la agencia	Hipótesis

Cuatro tipos de observación pide la ficha que los registros hechos por terceros no suelen traer: la ocasión (a qué responde la declaración); la audiencia (quién estaba presente y cómo reaccionó); la comparación del mismo hablante y del mismo evento ante audiencias diferentes, o antes y después de la verificación; y lo que ocurre después de la corrección (la forma permanece, varía o desaparece). Sin ellas, el análisis sigue siendo interpretación.

Ficha en blanco

Campo	Registro
1. Declaración	
2. Fecha	
3. Contexto y medio	
4. Audiencia presente	
5. Verificación (agencia, etiqueta, fecha)	
6. Operante probable	
7. Variables de control candidatas	
8. Marcas de forma	
9. Qué cambiaría la declaración y cómo se sabría	
10. Interpretaciones rivales y qué las separa	
No documentado	

APÉNDICE

B

Conceptos en uso

Cada concepto se presenta una vez, en un recuadro “Concepto en uso” (o “Definición propuesta”, cuando el término se propone aquí), en el primer capítulo cuyo análisis depende de él; los capítulos siguientes lo retoman y remiten al recuadro original. La tabla sigue el orden del libro.

Concepto	Capítulo (sección)	Definición breve	Fuente principal
Conducta verbal, audiencia y autoinforme	1 (1.5)	Conducta reforzada por la mediación de otras personas; la audiencia es el estímulo discriminativo en cuya presencia esta es característicamente reforzada. El autoinforme (escala, pregunta, entrevista) es conducta verbal, bajo control de quien pregunta y de cómo pregunta	Skinner (1957); Guerin (1994)
Creencia y actitud como conducta verbal	2 (2.8)	Las actitudes y las creencias son conducta moldeada por comunidades verbales, y no informes de un estado interno; la creencia puede refutarse con evidencia, la actitud no del mismo modo	Guerin (1994)
Autoclítico	2 (2.8); 3 (3.6)	Calificación añadida al enunciado que modifica su efecto sobre el oyente	Guerin (1994); Skinner (1957)
Triple contingencia, función × topografía y episodio verbal	3 (3.1)	El operante verbal se define por lo que controla la respuesta, y no por la forma; la unidad es ocasión, respuesta y consecuencia; el episodio incluye hablante y oyente	Skinner (1957)

Concepto	Capítulo (sección)	Definición breve	Fuente principal
Reforzamiento social generalizado	3 (3.3); retomado en 8 (8.1) y 9 (9.1)	Evento que suele preceder a muchos reforzadores diferentes y refuerza la conducta bajo varias condiciones de privación; la aprobación es el ejemplo. Para Guerin, las actitudes son enunciados mantenidos por él	Skinner (1957); Guerin (1994)
Tacto distorsionado y subcomunidad verbal	3 (3.4); retomado en 6 (6.1)	Tacto cuyo control por el evento fue desplazado por consecuencias especiales del oyente; una subcomunidad menor que la amplia puede controlar lo que se relata	Skinner (1957); Guerin (1992)
Autoclíticos de fuente y de certeza	3 (3.6)	Autoclíticos que informan la fuente o la fuerza de la respuesta; los cuantificadores (“todos”, “ninguno”) modifican la reacción del oyente a la oración entera	Skinner (1957)
Regla y conducta gobernada por reglas	3 (3.7)	Estímulo discriminativo verbal que describe una contingencia; el desempeño instruido es menos sensible a los cambios en las consecuencias	Baum (1995); Cerutti (1989); Hayes et al. (1986)
Actitud	4 (4.1)	Respuesta verbal de atribuir a un objeto un valor de positivo a negativo, mantenida por las consecuencias que dispone la comunidad verbal; no es necesario haber tenido contacto con el objeto	Guerin (1994); Cameoka y Moreira (2021)
Discriminación condicional e igualación a la muestra	4 (4.2)	Dados una muestra y unos estímulos de comparación, la elección se consecuencia según la muestra: si A1, entonces B1	Sidman y Tailby (1982)
Clase de equivalencia	4 (4.2)	Conjunto de estímulos cuyas relaciones condicionales tienen reflexividad, simetría y transitividad	Sidman y Tailby (1982)
Transferencia (o transformación) de función	4 (4.3)	Una función adquirida por contacto directo por un estímulo aparece en los demás miembros de la clase sin entrenamiento adicional; que se haya “transferido” es interpretación	Dougher et al. (1994); Dymond y Rehfeldt (2000)
Relaciones derivadas de oposición y de comparación	4 (4.6)	Patrones de responder relacional bajo control contextual: en la oposición, A opuesto a Y y a Z implica Y igual a Z; en la comparación, la relación define cómo se transforma la función	Dixon et al. (2003); Weinstein et al. (2008)
Contingencias conductuales entrelazadas	5 (5.1); retomadas en 9 (9.4)	La conducta de dos o más personas es evento ambiental para la de las otras; el resultado es el de la conducta interrelacionada	Glenn (2004)

Concepto	Capítulo (sección)	Definición breve	Fuente principal
Mando disfrazado	5 (5.2)	Enunciado con forma de tacto o de opinión cuya función, en la evidencia disponible, es la de mando; la aplicación al discurso político es una hipótesis propuesta aquí	Skinner (1957); Guerin (1994)
Control aversivo	5 (5.3)	Control de la conducta por estimulación aversiva (escape, evitación, castigo); puede ser verbal y leve	Skinner (1957); Fazzano y Gallo (2015)
Contracontrol	6 (6.4)	Episodios de condiciones controladoras aversivas socialmente mediadas y de respuestas de escape o evitación que no refuerzan la conducta controladora	Delprato (2002)
Evitación del contracontrol	6 (6.4)	Término propuesto aquí, como hipótesis: repertorio del hablante o de la comunidad mantenido por reducir la probabilidad o la eficacia del contracontrol	Propuesta hecha aquí, a partir de Delprato (2002)
Correspondencia decir-hacer	7 (7.4)	Relación entre lo que la persona dice que hizo o hará y lo que hace; son clases separadas, y la correspondencia es efecto de contingencia	Risley y Hart (1968); Lloyd (1994)
La atención como reforzador condicionado generalizado	8 (8.1)	La atención del oyente es consecuencia común de mandos generalizados y refuerza la conducta verbal sin importar la forma	Skinner (1957)
Reforzamiento intermitente	8 (8.4)	Solo parte de las respuestas o de las ocasiones va seguida del reforzador; en la conducta verbal, la intermitencia es la regla	Skinner (1957)
Programas de reforzamiento	8 (8.4)	El reforzador depende del tiempo transcurrido desde el anterior (intervalo) o del número de respuestas (razón), fijos o variables	Ferster y Skinner (1957)
Metacontingencia y macrocontingencia	9 (9.7); macrocontingencia retomada en 11 (11.4)	Metacontingencia: práctica cultural y resultado agregado de las variaciones en curso, que selecciona contingencias entrelazadas. Macrocontingencia: relación entre la práctica y la suma agregada de las consecuencias, que no controla al individuo	Glenn (1988, 2004); Houmanfar y Rodrigues (2006)
Audiencia negativa y audiencia no punitiva	10 (10.1)	Audiencia negativa: en su presencia la conducta verbal es castigada; no punitiva: solo no refuerza. Dos audiencias juntas generan un dilema	Skinner (1957)

Concepto	Capítulo (sección)	Definición breve	Fuente principal
Ventana de una audiencia	10 (10.1)	Definición propuesta: para un hablante y un asunto, los enunciados que emite ante esa audiencia sin que ella los castigue	Propuesta hecha aquí
Alcance, exposición y reacción	11 (11.2)	Definición propuesta: el alcance es quien recibe o puede ver la emisión; la exposición, lo que efectivamente vio; la reacción, lo que hace después	Propuesta hecha aquí
Entrenamiento en discriminación	12 (12.2)	El estímulo previo pasa a ser la ocasión de la respuesta por discriminación operante; con múltiples ejemplares, la respuesta queda bajo control de la propiedad común y se generaliza	Skinner (1957); Shea et al. (2023)
Diseño cultural	12 (12.7)	Intervenciones que cambian las contingencias operantes de muchos participantes o el ambiente que selecciona las prácticas (macroconducta y metacontingencias)	Glenn (2004)

APÉNDICE

C

Referencias

Las obras citadas en los capítulos, en orden alfabético (APA 7). Cuando autores y año coinciden, la letra que acompaña al año es la de las citas del texto. Todas se leyeron en texto completo, en la versión indicada en el Apéndice D (sección D.3).

Agência Lupa. (2026, 27 de julio). *Anatomia da desinformação eleitoral* [Informe]. 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), Artículo 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), Artículo 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), Artículo 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), Artículo 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), Artículo 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), Artículo 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), Artículo 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), Artículo 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), Artículo 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*. Publicación anticipada en línea. 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), Artículo 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), Artículo 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*, Artículo e024020. 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), Artículo 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), Artículo 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), Artículo 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 diciembre). *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 julio). *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. (Ed.). (2025a). *Formação e mudança de atitudes: Pareamento ao modelo e história em quadrinhos* (1.ª ed.). Instituto Walden4.

Moreira, M. B. (Ed.). (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* (1.ª 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., Morais, 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., Povoa, 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* (1.ª 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* (1.ª 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* (1.ª 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), Artículo 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* (1.ª 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), Artículo 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/ecomps.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

Ribeiro, A. (2023, 27 de diciembre). *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 diciembre). *Uma retrospectiva das mentiras do governo Bolsonaro* [Reportaje en cómic]. 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), Artículo 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), Artículo 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

APÉNDICE

D

Verificación de fuentes

NOTA DE PROCEDENCIA

La verificación registrada en este apéndice la hizo una inteligencia artificial (Claude, de Anthropic), que ayudó en la redacción del libro, y no los autores. Fue la IA la que armó la planilla de afirmaciones, obtuvo o recibió los textos de las fuentes por vía legal, buscó en ellos las frases que sostienen cada afirmación, leyó el contexto y registró los resultados y las salvedades.

Por eso, "leída" y "no leída", en las secciones D.2 a D.4, dicen lo que esa revisión logró leer y cotejar, y no lo que los autores leyeron o dejaron de leer. Una fuente "no leída" es una fuente que la revisión no pudo verificar: el libro no la trata como verificada y, donde el asunto exigía la fuente, el capítulo dice que no se leyó.

Este apéndice documenta cómo se cotejaron las afirmaciones de los capítulos con las fuentes, cuántas fuentes se leyeron y cuáles quedaron fuera, y qué decisiones de método afectan la lectura.

D.1 Cómo una afirmación llega a "verificada"

Cada afirmación atribuida a una fuente entra en una planilla con un identificador, la clave de la fuente, el capítulo, el enunciado y las frases literales de la fuente que la sostienen. Un script busca esas frases en el texto completo extraído (ignorando espacios, acentos, puntuación, separación silábica y ligaduras), registra la página y devuelve el contexto; quien decide es la lectura del contexto, que queda registrada con la fecha y las salvedades. Solo entonces la afirmación pasa a "verificada" y puede citarse. El texto de cada capítulo marca, en comentarios, las afirmaciones que sostienen cada párrafo, y un segundo script comprueba que toda cita del texto tiene referencia y que toda afirmación marcada existe y está verificada.

D.2 Números

El relevamiento registra 168 fuentes, de las cuales 108 se obtuvieron por vía legal y se leyeron en texto completo y 60 siguen pendientes (sección D.4). Se citan 124 obras. La planilla de afirmaciones tiene 1207 entradas, todas registradas como verificadas, con el contexto leído. La planilla completa (referencia, DOI, fragmento y página, estado) está en verificacao/tabela-de-verificacao.xlsx.

D.3 Versiones leídas que no son la de la revista

Cuando la versión leída no es la publicada, o no trae la paginación de la revista, el capítulo cita sin página y lo advierte. Las fuentes citadas en el libro en esa situación:

- Ajzenman et al. (2023): versión enviada, en RePEc.
- Álvarez-Benjumea y Winter (2018): versión enviada, en MPG.PuRe.
- Bail et al. (2018): texto en XML (PMC), sin la paginación de la revista.
- Brady et al. (2021): texto en XML (PMC), sin la paginación de la revista.
- Calleja et al. (2021): texto en XML (PMC), sin la paginación de la revista.
- Ceylan et al. (2023): texto en XML (PMC), sin la paginación de la revista.
- Cinelli et al. (2021): texto en XML (PMC), sin la paginación de la revista.
- Effron y Raj (2020): manuscrito aceptado, en London Business School Research Online.
- Eysenbach (2009): texto en XML (PMC), sin la paginación de la revista.
- Eysenbach (2020): texto en XML (PMC), sin la paginación de la revista.
- Ferster y Skinner (1957): edición original de 1957, Appleton-Century-Crofts, copia de PsycBooks del acervo del autor.
- Frimer et al. (2023): manuscrito aceptado, en PsyArXiv.
- Garrett y Bond (2021): texto en XML (PMC), sin la paginación de la revista.
- Guess et al. (2019): texto en XML (PMC), sin la paginación de la revista.
- Guess et al. (2020): texto en XML (PMC), sin la paginación de la revista.
- Hangartner et al. (2021): texto en XML (PMC), sin la paginación de la revista.
- Huszár et al. (2022): texto en XML (PMC), sin la paginación de la revista.
- Lasser et al. (2022): texto en XML (PMC), sin la paginación de la revista.
- Lazer et al. (2018): versión enviada, en arXiv.

- Mackinac Center (2019): página de internet, fuente interesada.
- Marino-Fages y Martínez-Pozo (2026): documento de trabajo, sin revisión por pares.
- Mota et al. (2023): versión enviada, en LA Referencia.
- Moxon et al. (1993): copia de ResearchGate (escaneo del PDF publicado).
- Nyhan y Reifler (2010): manuscrito del autor, sin la paginación de la revista.
- Rathje et al. (2021): versión enviada, en Apollo.
- Roozenbeek et al. (2022): versión enviada, en Apollo.
- Shea et al. (2023): texto en XML (PMC), sin la paginación de la revista.
- Siegel et al. (2021): versión puesta en línea por el autor.
- Skinner (1957): edición original de 1957, Appleton-Century-Crofts, copia del acervo del autor.
- Zielinski (2021): texto en XML (PMC), sin la paginación de la revista.

D.4 Fuentes no leídas

Fuentes previstas en el relevamiento para las que no se obtuvo el texto completo por vía legal. No se citan como verificadas; donde el asunto exigiría la fuente, el capítulo dice que no se leyó. De las 60 fuentes pendientes del relevamiento, 24 tienen un intento de obtención registrado en acceso abierto (repositorio, PMC, versión del autor) en verificacao/fontes.tsv; las demás no tenían acceso abierto localizado, y unas y otras dependen de que el autor adjunte el PDF. El último grupo de la lista son libros del propio autor, que quedan fuera de la excepción de las reglas de fuentes (sección D.5).

Acceso abierto, pero la descarga automática fue bloqueada (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), Artículo 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), Artículo 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), Artículo 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), Artículo 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), Artículo 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).

Sin acceso abierto 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).

Libros del autor fuera de la excepción (3)

- Efeitos reforçadores da atenção sobre o comportamento humano (libro del autor, fuera de la excepción de las reglas de fuentes) (capítulos 8, 9).
- Comportamento e práticas culturais (libro del autor, fuera de la excepción de las reglas de fuentes) (capítulos 9, 10).
- Modificação do comportamento verbal através de reforçamento em intervalo variável em situação análoga à psicoterapia: uma replicação sistemática (libro del autor, fuera de la excepción de las reglas de fuentes) (capítulo 9).

D.5 Decisiones de método que afectan la lectura

No se usan noticias como fuente de datos. La regla del proyecto para noticias (de los medios brasileños G1, CNN Brasil, BandNews y Canal Meio, con captura de pantalla, URL y fechas de publicación y de acceso) pide una recolección de páginas e imágenes que no pudo hacerse de modo auditable; los casos de declaraciones son los documentados en artículos e informes cuyo texto completo se leyó, citados con la página. La regla sigue valiendo para quien añada noticias.

Las declaraciones del Capítulo 3 son casos ilustrativos, y no una muestra. La muestra comparable del Capítulo 2 (sección 2.9) tiene una regla fijada y no se recolectó: Lupa y Aos Fatos (agencias brasileñas de verificación de datos) prohíben los rastreadores de inteligencia artificial en sus robots.txt, el feed público de ClaimReview cubre a Lupa de forma muy desigual y el informe de Lupa no distingue los mandatos de Lula. Antes de cualquier solicitud a un sitio, el robots.txt se leyó en un paso propio.

Dos libros entran por excepción, aprobada por el autor: Skinner (1957, *Verbal behavior*) y Ferster y Skinner (1957, *Schedules of reinforcement*), leídos en copias del acervo, con las páginas impresas. Los demás libros quedan fuera, salvo los del autor sobre actitudes y sobre cómics (historietas). Otros tres libros del autor, sobre los efectos reforzadores de la atención, sobre prácticas culturales y una replicación sistemática de modificación de la conducta verbal, no están en esa excepción: no se leyeron y no se citan, y reforzarían los Capítulos 8 a 10 si el autor amplía la excepción. El manual "O conceito de atitude na Análise do Comportamento", del autor, sigue inédito y no se cita; la definición de actitud en capas (Capítulo 4, sección 4.1) entra anclada en obras publicadas.

Los recortes de las imágenes del informe de Lupa del que el Capítulo 3 lee las clasificaciones están en verificacao/evidencias/, fuera del texto del libro.

D.6 Deslices de procedimiento registrados

Se registran dos deslices de método porque se afirma que el robots.txt se leyó antes de cualquier solicitud. El 11/10/2026, para obtener metadatos de Frimer et al. (2023), se envió una solicitud a api.osf.io antes de comprobar el robots.txt, que allí es Disallow: /; volvieron solo metadatos del resumen, que no se usaron, y el texto leído es el del PDF obtenido en osf.io/download. El 14/10/2026, al intentar obtener el texto de dos fuentes por PMC, se hizo una solicitud al efetch de eutils.ncbi.nlm.nih.gov en el mismo comando que impri-

mía el robots.txt, también Disallow: /; volvieron solo metadatos sin cuerpo, no se usó nada y los archivos se borraron. Desde entonces el robots.txt se lee en un paso propio, y no se usan las E-utilities del NCBI.

EDITORIAL

Editorial del Instituto Walden4

La Editorial del Instituto Walden4 se dedica a la promoción y difusión de conocimientos avanzados en la esfera del Análisis de la Conducta, así como en otras áreas prácticas basadas en evidencias, y abarca tanto aspectos científicos como profesionales. Comprometida con la democratización del acceso a la información, ofrece una amplia gama de libros de forma gratuita, y todos los títulos son accesibles en formato digital.

CONSEJO EDITORIAL

Dr. Gleidson Gabriel da Cruz

Dr. Márcio Borges Moreira

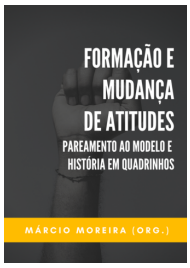
CONTACTO

secretaria@walden4.com.br

walden4.com.br · @instituto.walden4

CATÁLOGO

Otros libros de la Editorial del Instituto Walden4



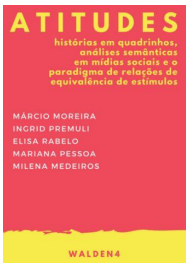
Formação e Mudança de Atitudes: Pareamento ao Modelo e História em Quadrinhos

Márcio Borges Moreira (ed.) · 2025



Preconceito: doze experimentos e um paradigma

Márcio Borges Moreira · 2022



Atitudes: histórias em quadrinhos, análises semânticas em mídias sociais e o paradigma de relações de equivalência de estímulos

Márcio Borges Moreira · 2022



Atitudes em relação a presidentes do Brasil: formação de classes de equivalência e transferência de função

Pedro Bufara Gonzaga Pereira, Márcio Borges Moreira · 2021



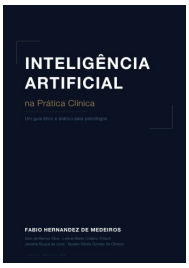
Efeitos reforçadores da atenção sobre o comportamento humano

Márcio Borges Moreira · 2021



Comportamento e Práticas Culturais

2015



Inteligência Artificial na Prática Clínica

Medeiros et al. · 2026



Do Encanto às Contingências

2026



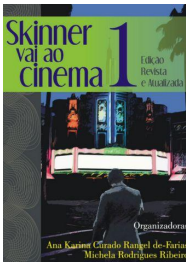
Uma história de aprendizagem operante

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

Conozca el catálogo completo (en portugués), con 65 títulos, muchos gratuitos

walden4.com.br/livros