

Desinformação Política, Mídias Digitais e Democracia: Como e Por Que as *Fake News* Funcionam?

Political Disinformation, Digital Media, and Democracy: Why and How Do Fake News Work?

JOÃO PAULO BACHUR¹

Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa (IDP). Brasília (DF). Brasil.

RESUMO: A desinformação é hoje objeto de intenso debate, principalmente por conta dos riscos à democracia. Os contornos do fenômeno, contudo, remanescem relativamente imprecisos. Desinformação é comumente associada à mentira ou à liberdade de expressão. O artigo sustenta que a desinformação deve ser entendida como operação social, e não como conduta individual: ela orienta o comportamento humano, a despeito da falsidade, ao permitir uma atribuição de sentido ao mundo. Como isso acontece? Para responder a essa pergunta, o trabalho integra quatro eixos de análise: (i) a fragmentação da esfera pública permite compreender o ambiente em que se desenvolve a ubiquidade das novas mídias digitais; (ii) o modelo de negócios baseado no engajamento *on-line* reverte características básicas da esfera pública: a integração de pontos de vista conflitantes; (iii) essa reversão produz efeitos para o indivíduo que simulam o comportamento de massa (contágio emocional e suspensão da racionalidade); e (iv) a reiteração de um enunciado linguístico produz um “efeito-verdade” para a desinformação ao permitir que o destinatário use o enunciado para atribuir sentido ao mundo. O artigo propõe uma sociologia interdisciplinar da desinformação, oferecendo uma descrição inédita e precisa dos processos sociais que estruturam o fenômeno.

PALAVRAS-CHAVE: Desinformação; mídias digitais; esfera pública; psicologia de massas; filosofia da linguagem.

ABSTRACT: Disinformation is currently a topic of intense discussion, mainly due to the risks it poses to democracy. The features of the phenomenon, however, remain relatively unclear. Disinformation is usually seen as a simple lie or as the exercise of the freedom of speech. The paper argues that disinformation should be understood as a social operation, and not as an individual conduct: despite its falsehood, it guides human behavior by ascribing meaning to the world. How does it work? To answer this question, this paper integrates four analytic levels: (i) the fragmentation of public sphere describes the environmental context in which the digital media reached ubiquity; (ii) the engagement-

1 Orcid: <https://orcid.org/0000-0002-3657-3965>.

based business model of on-line platforms reverses the key aspect of public sphere, namely the integration of conflicting points of view; (iii) this reversion effects individuals by simulating mass behavior (emotional contagion and suspension of rationality); and (iv) the reiteration of a linguistic statement produces a “truth-effect” in favor of the false information, for it enables the receiver to make sense of the world. The paper proposes an interdisciplinary sociology of disinformation, providing a unique and accurate description of the social process that structure the phenomenon.

KEYWORDS: Disinformation; digital media; public sphere; mass psychology; speech acts theory.

SUMÁRIO: Introdução; 1 Fragmentação da esfera pública: da credibilidade ao engajamento; 2 Modelo de negócios das plataformas digitais e polarização política; 3 Contágio emocional e efeito de massa pela mediação tecnológica; 4 Repetição, efeito-verdade, tomada de posição; Conclusões; Referências.

INTRODUÇÃO²

Este artigo investiga, em uma abordagem interdisciplinar, como o fenômeno da desinformação se estrutura *socialmente* – isto é, como a desinformação opera do ponto de vista da sociedade, e não do indivíduo isoladamente considerado. O impacto das chamadas *fake news* sobre processos democráticos, notadamente após Brexit, Trump e Bolsonaro, dentre outros, é hoje inquestionável (Benkler; Faris; Roberts, 2018; Benkler; Sunstein *et al.*, 2018; Vilmer *et al.*, 2018; ICO 2018, p. 47; Da Empoli, 2019, p. 12; Evangelista; Bruno 2019). Não obstante, as democracias ocidentais ainda não encontraram uma solução jurídica ou política para o problema: qual a fronteira entre desinformação e liberdade de expressão? Quais os limites da liberdade de expressão? Como evitar que tais limites se convertam em censura? E censura por parte de quem, das plataformas de comunicação digital? Podemos (ou devemos) colocar a democracia nas mãos das gigantescas empresas de comunicação digital? Temos alternativa? Todas essas questões são hoje objeto de intenso debate.

Talvez parte da dificuldade esteja relacionada a limitações na observação da desinformação como fenômeno social³. Muitos atribuem a disseminação de notícias falsas aos chamados *bots* (robôs), aplicações de *software* que simulam comportamento humano nas redes sociais replicando

2 Todas as citações feitas em português neste artigo são traduções livres do autor.

3 A pesquisa mais rigorosa tem evitado o uso do termo *fake news*, optando pelo conceito técnico de *desinformação*. Isso porque notórios propagadores de *fake news* empregaram esse termo contra a mídia tradicional, tornando a expressão politizada e disputável (Vosoughi *et al.*, 2018, p. 1146; Träsel, 2018, p. 71; Wardle; Derakhshan, 2017, p. 15).

conteúdos e interações (postagens) de forma automática. Contrariando essa intuição, criterioso levantamento empírico de pesquisadores do Media Lab (MIT) demonstra que robôs difundem notícias verdadeiras e falsas na mesma proporção: maiores volume e velocidade na disseminação de notícias falsas decorrem na verdade do comportamento humano, e não de *bots* (Vosoughi *et al.*, 2018). Por essa razão, os autores sugerem pesquisas qualitativas (entrevistas e até mesmo experimentos de neurociência) para aferir os fatores que interferem no “juízo humano” e favorecem a disseminação de notícias falsas (p. 1150).

Sem pretender invalidar essa rota para futuras pesquisas empíricas, a premissa que guia este trabalho é a de que a compreensão do fenômeno da desinformação não deve partir de comportamentos individuais, na tentativa de aferir se e por que (algumas) pessoas acreditam em notícias falsas, ou se e por que (algumas) pessoas compartilham notícias falsas conscientes da falsidade. A busca da *intenção subjetiva* nos parece pouco promissora por algumas razões, que apresentaremos de forma heurística e provisória, desenvolvendo-as ao longo deste estudo, quais sejam: (i) a desinformação é uma forma de comunicação socialmente estruturada, antecedendo e condicionando, por isso, o comportamento individual; (ii) a desinformação ativa gatilhos emocionais nem sempre acessíveis ao controle racional do indivíduo, pois que enraizados no inconsciente ou nas camadas mais profundas da psique; e (iii) a reiteração de enunciados linguísticos é determinante no processo de constituição de sentido operado pela desinformação, *i.e.*, na forma pela qual indivíduos atribuem sentido ao mundo.

Esses fatores relativizam a própria falsidade da desinformação ao permitirem que o indivíduo se oriente e se posicione frente a discussões éticas, políticas, científicas etc. Ou seja, o processo social pelo qual a desinformação operada torna *relativamente indiferente* a origem falsa (ou “mentirosa”) da informação, simulando algo como um “efeito-verdade” (Unkelbach *et al.*, 2019). Não se pretende sustentar que desinformação seja por isso indiferente: muito pelo contrário. O grande risco à democracia reside justamente no fato de que a desinformação permite que os indivíduos “entendam”, ou melhor, acreditem entender o mundo de uma maneira que lhes parece confiável, na qual a origem da falsidade é, por assim dizer, “apagada”.

Justamente por isso, soluções que apostam em agências de checagem ou códigos de conduta para o compartilhamento de memes e mensagens privadas soam ingênuas (Branco, 2018; Träsel, 2018).

A literatura especializada costuma mencionar que as notícias falsas não são um fenômeno novo, embora tenham sido potencializadas pelas mídias digitais. Diferença quantitativa, portanto (e não qualitativa). Parece haver aí uma equiparação entre desinformação e mentira. É claro, a mentira na política está longe de ser um fato novo. Basta recordar as reflexões de Hannah Arendt (1972) ou episódios recentes como a fala do então Secretário de Estado norte-americano Colin Powell perante o Conselho de Segurança da ONU, em fevereiro de 2003, quanto à existência de armas de destruição em massa no Iraque, suportando a guerra pretendida por George W. Bush; ou ainda os depoimentos de integrantes e ex-integrantes do governo Bolsonaro na CPI da Covid-19, entre maio e junho de 2021. A mentira é tão constitutiva da política quanto o risco de que a democracia se degenere em demagogia.

Este artigo sustenta que a desinformação caracteriza um fenômeno novo, distinto da mentira enquanto comportamento individual intencional e moralmente reprovável (cf. Simmel, [1899] 2016, e Derrida, 1996, para uma sociologia e uma filosofia da mentira, respectivamente). A desinformação pode ser compreendida como a

informação falsa ou enganosa que, cumulativamente: (a) é criada, veiculada e disseminada para obter ganho econômico ou intencionalmente enganar o público; (b) com potencial de causar dano público, entendido como ameaça a processos políticos democráticos, políticas públicas e bens públicos tais como a saúde dos cidadãos, o meio ambiente ou a segurança.⁴

Nessa definição, a desinformação parece ser um fenômeno qualitativamente diferente da mentira empregada de forma retórica ou estratégica no dia a dia da *Realpolitik*, e que opera socialmente mediante uma lógica específica.

Para esboçar a “lógica interna” (“*Eigengesetzlichkeit*”) da desinformação, este artigo apresenta quatro dimensões *sociais* do fenômeno: inicialmente, (i) apresentamos a fragmentação da esfera pública como contexto indispensável para a emergência da desinformação. Na sequência, veremos como (ii) o modelo de negócio das plataformas digitais se estruturou nesse

4 *Code of Practice on Disinformation* da União Europeia, documento de autorregulação assinado por Facebook, Google, Twitter, Mozilla e *stakeholders* do mercado publicitário em outubro de 2018. Disponível no *link*: <https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation>. Acesso em: 31.05.2021. Ver também Jack, 2017.

ambiente de fragmentação. Em seguida, passamos (iii) à análise dos efeitos emocionais e irracionais produzidos pelas novas mídias digitais, atualizando a tese canônica da psicologia de massas de Gustave Le Bon. Em último lugar, (iv) analisaremos como a desinformação funciona do ponto de vista de uma teoria dos atos de fala e da força performativa da linguagem. Veremos que o componente proposicional do enunciado linguístico (*i.e.*, sua correspondência a uma situação fática) é obnubilado pelo elemento performativo: o mundo bipolar que *a mensagem linguística cria, instando o destinatário a se posicionar*. Ao final, traremos algumas conclusões que podem iluminar a tarefa prática de regular, política e juridicamente, as plataformas digitais. Esse percurso interdisciplinar poderia ser descrito, de forma geral, como uma sociologia da desinformação. A desinformação, entendida como operação genuinamente social (*i.e.*, como operação da sociedade, e não como ato individual subjetivo), produz sentido e orienta o comportamento coletivo, *apesar* da falsidade que lhe caracteriza. Esse o problema a ser elucidado.

Neste artigo, focamos especialmente a desinformação política e sua relação com a democracia. O trabalho segue metodologia de análise bibliográfica interdisciplinar, cobrindo desde clássicos da sociologia e da psicologia social até recentes *reports* de *think tanks* especializados no tema⁵.

1 FRAGMENTAÇÃO DA ESFERA PÚBLICA: DA CREDIBILIDADE AO ENGAJAMENTO

O primeiro passo para avançarmos na descrição sociológica da desinformação é reconstruir o processo de fragmentação da esfera pública. Esse fenômeno pode ser descrito pela passagem de uma esfera pública estruturada pela *credibilidade da informação* a uma esfera pública (des)estruturada pelo *engajamento do usuário*. A reconstrução dessa passagem segue três etapas: (i) a formação da esfera pública literária e sua conversão em esfera pública *política*, historicamente marcada na passagem do século XVII ao século XVIII, culminando na Revolução Francesa; (ii) a consolidação de uma esfera pública de massas, que deu forma à indústria cultural entre meados do século XIX e meados do século XX; e (iii) a esfera pública digital.

Cada etapa e tipo de esfera pública tem meios de comunicação específicos como seu substrato material central: (i) a esfera pública política dos

5 Na temática de que se ocupa este artigo, é imprescindível recorrer a *reports* e *papers* elaborados por institutos de pesquisa, tais como, *e.g.*, Data & Society Research Institute.

séculos XVII e XVIII está associada à imprensa escrita e aos jornais diários; (ii) a esfera pública de massas, ao rádio e à televisão; e (iii) a esfera pública fragmentada, por sua vez, à internet, ao *smartphone* e às mídias digitais. A cada esfera pública corresponde um modelo jurídico de Estado: o Estado absoluto sucumbe com a ascensão da imprensa, abrindo caminho para o liberalismo. O Estado de bem-estar depende do consumo de massa e de uma mídia de massa. Após o neoliberalismo, o Estado contemporâneo vive imbricado nas redes comunicativas das mídias digitais (cf. Vesting, 2011a, 2011b, 2013, 2015).

Nenhuma outra teoria social está tão centralmente apoiada no conceito de esfera pública quanto a teoria do discurso de Habermas (1962). A esfera pública é responsável tanto por uma versão de democracia deliberativa baseada na busca pelo consenso quanto pelo restabelecimento das relações entre direito e moral nos termos de uma ética discursiva (Habermas, 1992).

A esfera pública pode ser definida como a reunião de pessoas privadas em um público (Habermas, 1962, p. 86). A definição parece tautológica, mas não é: acentua que a formação de um juízo político é feita por um público, em público. A investigação habermasiana parte do cânone historiográfico estabelecido por Koselleck em seu clássico *Crítica e crise* (1959). Koselleck reconstitui, a partir de Hobbes, o processo pelo qual a política se emancipa da moral e tem de se justificar por conta própria, sem recurso à legitimação teológica do poder absoluto – perante um *público*. Com o correr dos séculos XVII e XVIII, esfera pública passa a ser essa espécie de “tribunal moral” (Koselleck, 1959, p. 20 e ss.) que julga a pretensão de legitimidade do poder político. O processo pode parecer jurássico para usuários do Facebook, WhatsApp e Instagram, mas está longe de ser trivial.

A esfera pública política se forma a partir de uma esfera pública originalmente criada para o consumo de arte (Habermas, 1962, p. 116 ss.). O público se forma como público do teatro, das filarmônicas, dos museus: como o conjunto de pessoas privadas empenhadas em julgar a vanguarda estética do período. Esse processo, em que também a arte se emancipa da velha moral religiosa e se institucionaliza em organizações próprias, como museus, galerias, exposições, saraus, récitas etc. – a diferenciação funcional da arte (Luhmann, 1995) –, segue o desenvolvimento das cidades e, nelas, dos cafés e salões em que os burgueses se entretinham discutindo arte. É nesse contexto que os cidadãos desenvolvem uma habilidade essencial: a

capacidade de argumentar, em um debate público, a partir da ponderação de pontos de vista divergentes:

Museus, assim como concertos e peças de teatro, institucionalizam o julgamento leigo da arte: a discussão é o meio de apropriação estética. Os inúmeros panfletos de crítica ou apologia da teoria estética dominante se conectam com as conversas nos salões e são assim absorvidos – crítica estética como conversação. (Habermas, 1962, p. 102-103)

A discussão em torno de pontos de vista antagônicos – sobre os autorretratos de Rembrandt, os novos romances de Balzac ou Goethe, ou o uso do coro na nona sinfonia por Beethoven, por exemplo – institucionaliza a formação de juízos estéticos como processos *argumentativos públicos e coletivos* – o que não pressupõe consenso ou concordância, mas a contínua ponderação de argumentos divergentes. A esfera pública passa a ser o fórum de integração de pontos de vista conflitantes: a deliberação em público (inicialmente restrita à arte; depois expandida à política) mantinha pontos de vista conflitantes simultaneamente acessíveis a todos, isto é, não anulava qualquer dos pontos de vista em um consenso monolítico. Tratava-se de institucionalizar a *crítica* e o *debate* como operações sociais. Koselleck (1959) mostra que a antecâmara da Revolução Francesa foi processada de forma conspiratória, apoiada na maçonaria e em sociedades secretas. O que tornou a esfera pública revolucionária foi a transposição da crítica engendrada pelo segredo conspiratório para as luzes da esfera pública – que, *nesse processo*, tornou-se esfera pública propriamente *política*.

Com a Revolução Francesa e a consolidação das instituições parlamentares ao longo do século XIX, a esfera pública se institucionalizou como instância paralela aos parlamentos na formação da vontade política coletiva, apoiada na imprensa escrita. Essa a primeira transformação estrutural da esfera pública.

No século XX, a incorporação das massas trabalhadoras na esfera pública de origem burguesa e o correlato desenvolvimento da “indústria cultural” (Adorno; Horkheimer, 1947, p. 139/140) configuram uma nova mudança estrutural da esfera pública (Habermas, 1962, p. 326 e ss.). Os (então) novos meios de comunicação de massa, o rádio e a televisão, permitiram configurar o que se pode designar por esfera pública de *broadcasting*: transmissão de *uma* mensagem por vez, por veículo, para uma audiência indeterminada (Pariser, 2011, p. 22 e ss.). Como se trata de uma única mensagem endereçada a um destinatário potencial universal, a mensagem precisa ser

nivelada e “normalizada”, adotando-se aqui a imagem matemática de uma curva normal (*i.e.*, eliminando pontos de vista extremos), pois pressupõe uma apreensão relativamente homogênea (p. 128). Por essa razão, a mídia de *broadcasting* é muitas vezes caracterizada como um “intermediário de interesse geral” (Sunstein, 2017, p. 18): ela precisa representar a curva normal da opinião política, e a melhor forma para isso (com todas as imperfeições possíveis da mídia tradicional) é eliminar extremos.

A partir de meados dos anos 1990, tem lugar uma terceira mudança estrutural da esfera pública provocada pelos meios eletrônicos de comunicação. É com o uso massivo, em escala global, dos novos meios eletrônicos de comunicação (potencializado fundamentalmente pelos *smartphones*), que o processo de fragmentação da esfera pública se inicia. Niklas Luhmann, em um texto visionário – antes da fundação da Google e muito antes do Facebook, do WhatsApp ou do iPhone –, destaca o impacto dos meios eletrônicos de comunicação sobre a esfera pública, identificando como eles solapariam as bases em que se assentava a esfera pública tradicional:

Com efeito, a modificação que realmente traz consequências parece estar na invenção e no desenvolvimento de máquinas eletrônicas de processamento de informação. [...] Em muitos casos (exceção: telefone), a técnica impõe a unilateralidade da comunicação. Em parte, isso é consequência da intermediação por aparatos e, em parte, é uma exigência da comunicação de massas à qual a imprensa já se havia resignado. Isso altera os processos de seleção dos dois lados do aparato eletrônico. [...] O emissor escolhe temas e formas, encenações e sobretudo o momento da emissão levando em consideração o que lhe parece adequado. O receptor seleciona aquilo que quer ver ou ouvir. A comunicação ocorre então em um hiperciclo de seleções recíprocas, mas, a partir do momento em que ela ocorre, ela já não pode se autocorrigir. [...] Isso significa: a autoridade da fonte, com todas as suas salvaguardas sócio-estruturais imprescindíveis (estratificação e reputação) se torna dispensável, pois que anulada pela técnica e substituída por fontes anônimas e desconhecidas. (Luhmann, 1997, p. 303, 308 e 309)

A esfera pública tradicional, apoiada no jornalismo de *broadcasting*, era estratificada e organizada pelos grandes conglomerados jornalísticos. Essa estratificação provocava, por óbvio, assimetrias e hegemonias de caráter político (Bourdieu, 1994). Não obstante, ela exercia uma função para a sociedade como um todo: provia *parâmetros interpretativos* com recurso aos quais os indivíduos podiam interpretar a representação do mundo fornecida pela grande mídia. “Tudo que nós sabemos sobre nossa sociedade,

e também sobre o mundo em que vivemos, sabemos através dos meios de comunicação de massa” (Luhmann, [1995b] 2004, p. 9). Ao lermos uma determinada notícia, temos acesso ao produto final de uma cadeia de editoração e edição jornalística, guiada (pelo menos em tese) pelo princípio da apuração imparcial e pelo compromisso (pelo menos formal) com a verdade dos fatos⁶. Somos então capazes de avaliar a cobertura jornalística de um determinado evento, julgando-a conforme critérios de credibilidade, reputação (do veículo, das fontes etc.) e confiabilidade. É claro que essa estrutura piramidal da esfera pública tradicional não poderia ser designada como democrática. Apesar disso, aquelas “salvaguardas socioestruturais imprescindíveis (estratificação e reputação)” de que fala Luhmann disponibilizavam recursos para que os indivíduos pudessem interpretar o mundo retratado pelos meios de comunicação de massa, concordando ou discordando da mensagem veiculada – apontando, inclusive, eventual parcialidade. No caso brasileiro, há bem pouco tempo era possível discordar ou criticar, mas nunca simplesmente ignorar o que aparecia, digamos, no Jornal Nacional. Ninguém suporia que, em um intervalo muito curto de tempo, vídeos produzidos para canais do YouTube e replicados em grupos de WhatsApp ou Telegram poderiam disputar o monopólio jornalístico da televisão de sinal aberto⁷.

Essa credibilidade da mídia tradicional foi erodida, em medida significativa, pelos novos meios digitais de comunicação. As novas mídias digitais tornaram a internet acessível a qualquer um – e, se pode haver aí algum ganho de democratização (o que questionaremos na próxima seção), a produção, circulação e replicação infinitesimal de mensagens e conteúdos produzidos por indivíduos privados imerge os usuários em um oceano de ruído. Esse ruído precisa ser apreensível para fazer (algum) sentido. É preciso reduzir complexidade. O capitalismo das plataformas digitais é então caracterizado por uma disputa pela *atenção do usuário*. A torrente de postagens, compartilhamentos e interações geradas pelos usuários das novas mídias digitais exigiu algum mecanismo de filtro – os algoritmos – que se aprimoraram para captar aquilo que potencialmente pudesse interessar a

6 É claro que isso não nega a existência de coberturas parciais. Não obstante, são esses frágeis princípios que nos permitem identificar, inclusive, desvios na cobertura jornalística da mídia tradicional.

7 Um exemplo: o canal de YouTube *Foco do Brasil* tem quase três milhões de seguidores e se apresenta como veículo para “As notícias e ações do Presidente Jair Bolsonaro e seu governo e o nosso Brasil” (cf. *link*: <https://www.youtube.com/c/FocodoBrasil>). Trata-se de importante fonte de vídeos bolsonaristas que circulam largamente em grupos de aplicativos de mensagens privadas.

cada usuário. Logo passaremos a isso em maior detalhe. O ponto é que, com Luhmann, vê-se que “a técnica anula a autoridade da fonte e a substitui pelo irreconhecível da fonte”.

Da autoridade da fonte à fonte anônima; da credibilidade ao engajamento: as transformações da esfera pública (com a mídia impressa nos séculos XVIII e XIX; o *broadcasting* de massa no século XX e as novas mídias digitais de hoje) estão, na verdade, inseridas em um longo processo de erosão dos aparatos hermenêuticos que sempre nos ajudaram a atribuir sentido ao mundo. A melhor referência aqui continua sendo Luhmann (1997, p. 145-324): esse processo, que culmina nos meios eletrônicos de comunicação, está relacionado à fratura do monopólio hermenêutico da igreja católica provocada pela Reforma Protestante e indissociável tanto da tradução da Bíblia quanto da difusão da imprensa. Aqui precisamos ser esquemáticos para não comprometer o fluxo do argumento: a tradução da Bíblia para o alemão e o inglês abriram caminho para a crítica leiga do texto fundante do Cristianismo, não sendo mais possível assegurar *uma única* interpretação *unitária* do mundo. Com isso, o passo para questionar o Antigo Regime e o direito divino dos reis era questão de tempo. Os séculos XIX e XX consolidaram o *broadcasting*, que passou a ser um dos principais aparatos hermenêuticos para nossa compreensão do mundo. As mídias digitais fragmentam esses aparatos. Os usuários das novas mídias digitais, via de regra, não conseguem aferir se aquilo que leem em seu celular é representativo dos principais problemas sociais e políticos do momento, se as fontes são fidedignas e se, é claro, não se trata de desinformação. A perda desse crivo de comparabilidade, proporcionalidade e credibilidade, promovida pelas novas mídias sociais, é precisamente o que caracteriza a fragmentação da esfera pública. As consequências da fragmentação da esfera pública, no que diz respeito aos riscos à democracia, estão relacionadas à formação das bolhas (“*filter bubbles*”) e câmaras de ressonância (“*echo chambers*”), bem como à polarização política que a interação nesse contexto digital provoca.

O ganho de importância das redes sociais resulta, ademais, em uma *ascensão de particularismos* e no *retrocesso* de uma parte da esfera pública a *fóruns de pessoas que pensam da mesma forma*. A formação de tais fóruns é reforçada também pela inserção de algoritmos de aprendizagem automática que, de forma direcionada, recompensam contribuições específicas que desencadeiam fortes emoções e interações diretas e terminam por ter como resultado “criar para o usuário individual seu mundo próprio e singular”. Aqui se chega a uma autolimitação temática preocupante, a uma *cegueira narcisista em relação às realidades da vida* que se encontram fora do pró-

prio grupo e que *não permitem mais uma percepção mútua de imagens de mundo e visões de mundo de outros milieus culturais*". (Vesting, 2018, p. 102/103 – grifos acrescentados)

Nesse passo, é preciso descrever o modelo de negócio das plataformas e relacioná-lo à polarização política.

2 MODELO DE NEGÓCIOS DAS PLATAFORMAS DIGITAIS E POLARIZAÇÃO POLÍTICA

Como adiantamos na introdução, a desinformação “funciona” socialmente ao permitir que os indivíduos se orientem politicamente. Para que essa orientação seja possível, é indispensável que o destinatário das *fake news* perceba o mundo em uma chave bipolar (Barfar, 2019). Todo e qualquer tema político deve ser apresentado em uma versão “contra/a favor”, oferecendo a opção de se situar em um dos lados. A desinformação, especialmente na política, exige polarização. Nesta seção, veremos como o aprimoramento dos algoritmos das mídias digitais aprofundou a polarização política, ameaçando a democracia⁸.

Claro, há visões que questionam a associação entre as novas mídias digitais, desinformação e polarização política:

Não há evidências concretas de que a difusão de *fake news* mude o curso de eleições, nem que o aumento de *bots* impacte negativamente como as pessoas discutem política. Não vivemos em um mundo pós-verdade. [...] Há problemas com a busca do Google, os filtros do Facebook e o *trending* do Twitter. Mas precisamos lembrar que essas ferramentas são excepcionais. (Sumpter, 2018, p. 172 e 175)

Exposição à discordância política em mídias sociais parece elevada, acesso à internet e uso de mídias sociais não estão correlacionados com crescimento da polarização e a má informação [*“misinformation”*] parece ter apenas efeitos limitados no nível de conhecimento político dos cidadãos. (Tucker *et al.*, 2018, p. 15)

Ora, ninguém questiona o valor das plataformas digitais para os nossos dias. Hoje é praticamente impossível realizar algumas das tarefas mais

8 No que segue, valemo-nos essencialmente das descrições de Pariser, 2011; O’Neil, 2016; Srnicek, 2017; Sumpter, 2018; Da Empoli, 2019; e Applebaum, 2020. Embora cresçam as publicações acadêmicas sobre as plataformas, as principais obras têm ainda o caráter de *Sachbuch* (em parte devido à escassez de informações sobre o funcionamento preciso dos algoritmos desenvolvidos pelas plataformas; em parte porque os principais financiadores de *think tanks* são as próprias *Big Techs*).

comezinhas da vida cotidiana sem os recursos disponibilizados pelas novas mídias digitais. A continuidade de algumas atividades econômicas, acadêmicas e profissionais durante a pandemia de Covid-19 só foi possível graças a essas tecnologias (mas a tecnologia também intensificou, justamente no contexto da pandemia, o monitoramento e a vigilância dos cidadãos, especialmente em países como a China, para ficarmos com o exemplo mais notório). As novas mídias digitais alteraram a sociabilidade de maneira profunda, estrutural e permanente. Há vinte anos ninguém se expunha da forma permitida hoje por Facebook, Twitter ou Instagram; ninguém reagia a tudo que circula nas redes sociais de forma praticamente instantânea. Essa nova sociabilidade é possível graças ao substrato técnico que altera a forma como as pessoas interagem e percebem o mundo. E essa nova sociabilidade muda o comportamento – *on-line e off-line* – o que só torna mais importante investigar os efeitos potencialmente disfuncionais dessas novas tecnologias.

Não pretendemos, por isso, imputar às novas mídias sociais uma espécie de causalidade mecânica na produção da polarização política. O que pretendemos acentuar é que essas novas tecnologias adquiriram funcionalidades que potencializaram o impacto e o alcance da desinformação. Se as plataformas não criaram a polarização, seus algoritmos também não a bloqueiam. Ao contrário, intensificam-na. O fator tempo é crucial: por óbvio, não é a criação das plataformas digitais *per se* que compromete a democracia. O Facebook foi de fato criado para integrar comunidades acadêmicas (quase como um *site* de encontros muito sutil). O buscador da Google foi criado para entregar ao usuário o melhor resultado para a busca pretendida. A Amazon busca de fato sugerir livros que possam nos interessar. Mas, com o tempo, cada vez mais pessoas passaram a dedicar cada vez mais horas de seus dias à interação nas novas mídias digitais. Com o tempo, essas plataformas passaram a dispor de volume inédito de dados pessoais. Sem alguma forma de filtro ou priorização, a tecnologia se anularia. Os filtros de relevância surgem como uma solução técnica para organizar a informação da plataforma para o próprio usuário⁹. Por isso, algoritmos de relevância associados ao *targeted-marketing* produziram efeitos de médio prazo potencialmente disfuncionais.

9 O exemplo mais notório é o algoritmo *EdgeRank*, que estrutura o *News Feed* na página inicial do Facebook: “At first, the News Feed showed nearly everything your friends did on the site. But as the volume of posts and friends increased, the Feed became unreadable and unmanageable. Even if you had only a hundred friends, it was too much to read. Facebook’s solution was *EdgeRank*, the algorithm that powers the default page on the site, the *Top News Feed*” (Pariser, 2011, p. 36).

Como vimos, as novas mídias digitais provocaram uma transformação essencial na esfera pública: a tradicional mídia de *broadcasting* perdeu o monopólio como ferramenta de interpretação do mundo. É claro, a relação entre a mídia tradicional e as novas mídias é complexa e objeto de intenso debate. Se, por um lado, as novas mídias digitais passaram a ser empregadas como fonte predominante de informação, sabe-se que muito do que circula nas redes sociais é originado pela mídia tradicional (Tucker *et al.*, 2018)¹⁰. Não obstante, as novas mídias digitais mudaram a forma como as pessoas percebem o mundo e suas relações interpessoais. Nesse aspecto específico, elas tomaram o lugar da tradicional mídia de *broadcasting*. Mesmo a mídia tradicional é hoje apreendida (*i.e.*, consumida de forma massiva) por intermédio de redes sociais e aplicativos de smartphones. O crescimento das mídias digitais coincide com declínio da mídia tradicional (Wardle; Derakhshan, 2017, p. 51). E o fio condutor para operar essa mudança parecer ter sido uma profunda alteração na lógica publicitária.

Anunciantes perceberam que o *marketing* na mídia de *broadcasting* era muito menos efetivo e muito mais caro do que o *marketing* apoiado na coleta intensiva de dados pessoais, *profiling* e *micro-targeting*: ao invés de investir em anúncios de página inteira em jornais diários ou em inserções no horário nobre da televisão, tornou-se mais certo e muito mais barato apostar no *targeted-marketing*, isto é, na mensagem publicitária especificamente desenhada para um público pré-determinado (Pariser, 2011, p. 47/49). Essa inflexão é decisiva para a consolidação do modelo de negócios das plataformas digitais. Em um primeiro momento, ela foi viabilizada pelo uso dos algoritmos nas plataformas; na sequência, passou a motivar o aprimoramento e a intensificação no emprego desses algoritmos. O desenho das interfaces digitais e das funcionalidades das plataformas tem um efeito direto na fragmentação da esfera pública em função das chamadas *bolhas* (“*filter-bubbles*”) ou *câmaras de ressonância* (“*echo chambers*”), espaços digitais em que o usuário vê sua própria imagem replicada e ampliada – o que, por sua vez, amplifica e intensifica a polarização política.

O momento decisivo em que as plataformas digitais transformaram a economia mundial de fins do século XX e início do XXI está em associar uma determinada atividade empresarial à *personalização* dessa atividade.

10 Há países que regulamentaram a remuneração da mídia tradicional como compensação pela circulação do conteúdo jornalístico, como Alemanha, Espanha e mais, recentemente, a Austrália (é o chamado *link tax*). A medida é polêmica e marcada por uma série de efeitos paradoxais (Athey *et al.*, 2017).

Isso significa que cada consumidor/usuário passa a ter uma experiência única em uma determinada plataforma. Ao ser fundada em 1994, a Amazon pretendia replicar, na compra de livros *on-line*, a experiência reconfortante das pequenas livrarias de bairro, em que o dono recomendava sugestões de leitura a seus clientes habituais (Pariser, 2011, p. 28). Para isso, empregava a Tapestry, ferramenta desenvolvida pelo Xerox Palo Alto Research Center, *think tank* responsável por inovações tecnológicas adotadas mundialmente, como o *mouse* ou a interface de usuário dos computadores (*graphical user interface*: a tela com ícones e um cursor para apontar, controlado pelo *mouse*). Tapestry foi o primeiro *filtro colaborativo* desenhado para identificar relevância: a ferramenta organizava as caixas de entrada de *e-mails* a partir do histórico do comportamento dos usuários, ordenando as mensagens potencialmente relevantes. A Amazon revolucionou o *e-commerce* ao empregar filtros de relevância para quebrar o caráter impessoal da compra *on-line*. “E, é claro, você não precisava usar [filtros de relevância] só para *e-mails*” (Pariser, 2011, p. 27).

Filtros colaborativos poderiam ser utilizados para qualquer fluxo de informação grande o suficiente para exigir uma ordenação por relevância. A relevância é definida pelo histórico do usuário em lidar com esse fluxo de informação. E se, em fins dos anos 1990, apenas a Amazon utilizava a tecnologia, tentando acertar na recomendação de livros para seus consumidores, esse tipo de filtro passou a ser indispensável à medida que o volume de informação e transmissão de dados na internet adquiriu escala planetária – o que se tornou realidade com os *smartphones*. A partir de então, quando bilhões de pessoas passaram a gerar milhões de dados por minuto, esse tipo de filtro colaborativo para indicar relevância passou a ser a tecnologia central – tanto para rastrear o percurso de navegação *on-line* por usuários do buscador Google quanto para organizar o *feed* de notícias e notificações do Facebook, do Twitter ou do Instagram, por exemplo. Nisso reside o paradoxo das novas mídias digitais: a chave para identificar a informação relevante está em processar uma quantidade muito grande de dados, mas essa quantidade de dados exige filtros cada vez mais refinados – o que exige, por sua vez, cada vez mais dados.

Essa tecnologia de filtros de relevância com base em dados permitiu customizar a experiência virtual de cada usuário, tornando sua experiência *on-line* única. Trata-se do “*daily me*”, a imagem antevista por Negroponte, um dos fundadores do Media Lab do MIT: “Com o Daily Me [...] você não vai mais confiar ao jornal local a curadoria do que você vê, e você poderá

contornar as redes de televisão. Ao contrário, você poderá desenhar um pacote de comunicação só para você, com cada componente escolhido de antemão” (citado por Sunstein, 2017, p. 1). Esse prognóstico de meados dos anos 1990 acabou por se materializar no fenômeno das *bolhas*:

Os algoritmos que orquestram os anúncios que vemos começaram a orquestrar nossas vidas. O código básico no coração da nova internet é bem simples. A nova geração de filtros observa as coisas que você parece gostar – as coisas reais que você tem feito ou as coisas que pessoas parecidas com você gostam – e tenta extrapolar. São máquinas de predição, constantemente criando e refinando uma teoria de quem você é e do que você vai fazer ou desejar a seguir. Juntas, essas máquinas criam um universo de informações único para cada um de nós – o que eu chamei de filtro-bolha – que altera profundamente a forma como encontramos ideias e informações. (Pariser, 2011, p. 8)

Há, claro, um quê de dramaticidade nessa descrição. Afóra o exagero nas cores (“algoritmos começaram a orquestrar nossas vidas”), o importante a reter é que as bolhas produziram algo como uma *perda de perspectiva*: se a mídia de *broadcasting* permitia uma visão geral (digamos, uma janela, de onde se observava a aleatoriedade da rua), as bolhas oferecem um olhar afunilado, fragmentado e parcial; a visão ampliada de um recorte minúsculo da realidade (um microscópio). A esfera pública tradicional está relacionada à cidade e à aleatoriedade da informação: ao folhear um jornal ou assistir à TV de sinal aberto, recebemos a mensagem que nos é transmitida – tal como quando caminhamos pelas ruas de uma cidade (Sunstein, 2017, p. 12) – o que nos obriga a interagir com pontos de vistas divergentes. “Quando as pessoas usam o Facebook para verem exatamente aquilo que querem ver, a sua compreensão do mundo é profundamente afetada” (p. 2). Quando passamos a maior parte do tempo interagindo em nossas bolhas virtuais, perdemos a dimensão do todo. E o problema é que o desenho das novas mídias sociais nos impulsiona justamente para isso.

As grandes empresas de tecnologia são, antes de tudo, *empresas*: organizações comerciais que, para maximizar seus rendimentos, precisam encorajar os usuários a permanecerem *on-line* o maior tempo possível, ampliando a exposição ao *marketing* (Wardle; Derakhshan, 2017, p. 52). Isso é feito com base no engajamento pretérito do usuário (curtidas, comentários, compartilhamentos etc.). Por isso a dimensão temporal é importante: tanto em relação ao tempo que cada usuário gasta interagindo nas novas mídias digitais quanto em relação ao tempo que as plataformas tiveram para

acumular dados pessoais de seus usuários. Para as plataformas, o ideal é que todos gastem o maior tempo possível *on-line*. O desenho dos algoritmos buscou então justamente isso. Especialmente, por exemplo, no caso do Facebook:

Nós fornecemos a você uma pequena dose de dopamina cada vez que alguém o curte, comenta uma foto ou um *post*, ou qualquer outra coisa sua. É um *loop* de validação social, exatamente o tipo de coisa que um *hacker* como eu poderia explorar, porque tira proveito de um ponto fraco da psicologia humana. Os inventores, os criadores, eu, Mark [Zuckerberg], Kevin Systrom, do Instagram, estávamos perfeitamente conscientes disso. E, mesmo assim, fizemos o que fizemos. *E isso transforma literalmente as relações que as pessoas têm entre si e com a sociedade como um todo*. Interfere provavelmente na produtividade, de certa maneira. Só deus sabe qual o efeito que isso produz nos cérebros de nossos filhos. (Sean Parker, primeiro financiador do Facebook, *apud* Da Empoli, 2019, p. 52/53 – grifos meus)

A dopamina de que fala Sean Parker é um neurotransmissor que atua regulando o comportamento motivado por recompensa. E, como se sabe, está associada a diferentes tipos de vícios. Cada *like* que recebemos dispara uma microdose de dopamina e nos motiva a voltar às redes sociais (Blachnio *et al.*, 2016). Essa nova sociabilidade digital mediada pelas plataformas permite que algoritmos psicométricos reconstituam a personalidade dos usuários (*profiling*) e, para maximizar as chances de interação, montem cada *feed* individual com postagens potencialmente mais capazes de desencadear novas curtidas, novos compartilhamentos etc., agregando *clusters* de usuários que pensam de forma semelhante. Essa autorreferência motivada pela personalização da experiência *online* (indispensável para o *targeted-marketing*) faz as bolhas digitais funcionarem *limitando* a percepção de mundo dos usuários por aquilo que a psicologia designa como *viés de confirmação*: a tendência de acreditar naquilo que confirma nossa visão de mundo pré-existente. “Os filtros-bolha tendem a amplificar dramaticamente o viés de confirmação – em certo sentido, eles foram desenhados justamente para isso” (Pariser, 2011, p. 88). E as consequências para o incremento da polarização política são diretas: “O problema político mais sério apresentado pelos filtros-bolha é que eles tornam cada vez mais difícil a ocorrência do debate público” (p. 155). Insulados em bolhas que refletem apenas múltiplas visões de nossa própria opinião, nossos valores e princípios pessoais, nossa opinião política e a indubitabilidade de nossas convicções são amplificados e radicalizados. O resultado é a polarização política (Sunstein, 2017, p. 10).

Como isso ocorre? Diversos experimentos da psicologia de grupo demonstraram que a discussão com pessoas que pensam de maneira semelhante sobre determinados temas tende a reforçar os pontos de vistas individuais pré-existentes à discussão (Sunstein, 2009). Basicamente, os experimentos dividem grupos focais e controle, compostos com pessoas que pensam de maneira semelhante em determinados temas (e.g., aborto, pena de morte etc.), e pessoas que pensam de forma divergente. Há, portanto, grupos homogêneos contrários e a favor do aborto, por exemplo, e grupos heterogêneos compostos por pessoas a favor e pessoas contrárias ao aborto. Os experimentos indicam que, nos grupos homogêneos (independentemente da posição, *i.e.*, tanto para os que são contrários quanto para os que são favoráveis ao aborto), os participantes saem das discussões *tendo radicalizado seu próprio posicionamento pré-existente*. Isso ocorre porque: (i) o pequeno grupo tem apenas uma amostra limitada de argumentos, não desenvolvendo contra-argumentos; (ii) não há incentivos para mudar a opinião do grupo por parte daqueles que detêm posições menos extremas ou radicais; (iii) a legitimação por pares sugere que a própria posição é “a correta”; e (iv) o pequeno grupo ativa mecanismos de identificação e pertencimento que muitas vezes se impõem às posições pessoais (ou seja, é possível ocultar aquilo que se pensa em nome de preservar uma identidade coletiva) (Sunstein, 2017, p. 74; Talisse, 2019, p. 102).

Em relação ao posicionamento político, a polarização assume contornos especialmente preocupantes, pois favorece a emergência de posições extremadas (Benkler; Faris; Roberts, 2018). A “discussão em grupos de pessoas que pensam da mesma forma produz uma polarização relativa a crenças” (“*belief polarization*” – Talisse, 2019, p. 105). Se a projeção política da esfera pública de *broadcasting* poderia ser vista no que a ciência política designa como teorema do eleitor mediano (uma curva normal no espectro político ideológico esquerda-direita, concentrando a maioria do eleitorado no centro), a esfera pública fragmentada das plataformas digitais permite a proliferação de opiniões políticas menos sensíveis a ouvir qualquer contraponto: “Nós trocamos intensamente mensagens políticas clamando por indignação, ultraje, ira e apoio apenas entre aqueles que já comungam de nossa perspectiva política” (Talisse, 2019, p. 6). Protegido no pequeno grupo, isolado em sua bolha, o indivíduo passa por um processo de des-individualização (“*deindividuation*” – Festinger; Pepitone; Newcomb, 1952) e se torna parte de um fenômeno coletivo. Essa transubstanciação do indivíduo em massa é há muito estudada pela sociologia e pela psicologia, e tem um componente afetivo ou emocional muito claro. A dimensão inovadora

– sem a qual o funcionamento social da desinformação não fica claro – é a simulação do efeito de massa para o indivíduo sem a aglomeração física, por força da mediação tecnológica.

3 CONTÁGIO EMOCIONAL E EFEITO DE MASSA PELA MEDIAÇÃO TECNOLÓGICA

As novas tecnologias da comunicação produzem um efeito paradoxal na nova sociabilidade por elas engendrada: mesmo estando sozinhos, encapsulados nas bolhas digitais que rodam em nossos *smartphones*, estamos simultaneamente mais vulneráveis aos efeitos de massa. Aquilo que a psicologia social chama de efeito de massa – a suspensão da racionalidade individual por força de um contágio emocional – ocorre agora mediado pela tecnologia e independente de uma multidão concreta.

A psicologia das massas foi inaugurada como disciplina por Gustave Le Bon (1895). Le Bon foi o primeiro a identificar como a ação inconsciente do coletivo se impõe de forma poderosa, mas momentânea, sobre a atividade individual consciente. Esse contágio emocional ocorre por *sugestão*: a adoção inconsciente de um padrão de comportamento formado pela agregação física de indivíduos em um determinado momento, em uma determinada circunstância (Le Bon, 1895, p. 20)¹¹. As características do comportamento de massa são conhecidas: a massa é “impulsiva, volúvel, excitável”, “extraordinariamente influenciável e crédula, é acrítica, o improvável não existe para ela” (Freud, [1921] 2011, p. 25); nela ocorre o “enfraquecimento da aptidão intelectual, a desinibição da afetividade, a incapacidade de moderação e adiamento, a tendência a ultrapassar todas as barreiras na expressão de sentimentos e a descarregá-los inteiramente na ação” (p. 77). “Na massa o indivíduo está sujeito a condições que lhe permitem se livrar das repressões dos seus impulsos instintivos inconscientes” (p. 21). Em síntese, trata-se de um estado de regressão em que a afetividade se transmite entre os indivíduos e bloqueia o uso consciente da razão:

[...] a atrofia da personalidade individual consciente, a orientação de pensamentos e sentimentos nas mesmas direções, o predomínio da afetividade e da psique inconsciente, a tendência à imediata execução dos propósitos que surgem – tudo isso corresponde a um estado de regressão a uma atividade anímica primitiva. (Freud, [1921] 2011, p. 85)

11 Freud critica o poder explicativo limitado do conceito de “sugestão” de Le Bon, acrescentando uma variável libidinal – mas não entraremos nesses pormenores aqui. Para os fins deste artigo, a sugestiva noção de *contágio emocional* é suficiente (Hatfield; Cacioppo; Rapson, 1994).

Em síntese, o efeito de massa designa esse predomínio da emoção sobre a racionalidade individual, com a correlata desinibição do comportamento instintivo e pré-consciente. Por isso Freud descreve o efeito de massa como uma *regressão*: como se o superego que o indivíduo forma à medida que passa da infância à juventude e à idade adulta fosse momentaneamente solapado pelo instinto primitivo e imediatista. O fenômeno depende do ajuntamento físico de pessoas, de uma sincronização emocional quase espontânea – quase espontânea porque a primeira faísca é sempre produzida por alguém e esse primeiro movimento pode não ser espontâneo. Mas daí em diante, o contágio emocional se espalha. Elias Canetti designa esse fenômeno coletivo fortemente emotivo como *descarga* (“*Entladung*”): uma liberação emocional – uma espécie de descarrego – pela qual todos os membros da massa se sentem iguais (Canetti, [1960] 1995, p. 22). Há nessa descarga emocional um importante componente sociológico de *integração social*, de formação de laços e de pertencimento a um grupo. O isolamento social é superado em nome da formação de um coletivo. E a afetividade que guia esse fenômeno é arrebatadora.

Como imaginar que todos esses fenômenos podem vir a ocorrer sem a formação física da massa?

Em sua clássica investigação sobre a democracia, Schumpeter trata da irracionalidade política que emerge com a massa. O pensamento político do homem médio “assume o caráter puramente associativo e afetivo” (idem, p. 319), suspendendo os “freios morais e maneiras civilizadas de pensamento e sentimento” e permitindo, ao mesmo tempo, a “erupção de impulsos primitivos, infantilismo e tendências criminais” (1942, p. 320). Até aqui, Schumpeter segue Le Bon e Freud. O ponto a ressaltar é o seguinte: a formação da massa pode ocorrer pela mediação tecnológica, ou seja, sem a presença física de uma multidão:

Esses fenômenos *não estão limitados à multidão no sentido de aglomeração física de numerosas pessoas. Leitores de jornal, audiências de rádio, membros de partidos políticos, mesmo quando não fisicamente reunidos, podem ser facilmente transformados psicologicamente em multidão e levados a um estado de frenesi, no qual qualquer tentativa de se apresentar um argumento racional desperta apenas instintos animais. [...] a mera afirmação, repetida constantemente, vale mais do que o argumento racional.* (Schumpeter, 1942, p. 313/314 – grifos nossos)

O próprio Le Bon já havia antecipado essa circunstância ao tratar o efeito que a leitura de notícias poderia desencadear: “O contágio não exige

a presença simultânea dos indivíduos em um determinado local; ele pode se operar à distância” (Le Bon, 1895, p. 67). Ora, a terceira mudança estrutural da esfera pública (a passagem da mídia de *broadcasting* para a bolha das redes sociais) torna as tendências prenunciadas por Le Bon e Schumpeter uma realidade: ao criarem bolhas de interação em que a própria opinião ecoa como verdade indisputada e ampliada, os algoritmos *simulam* uma multidão *tailor-made* para cada usuário, permitindo que o frenesi da emoção coletiva se propague via *smartphone*, em cada indivíduo isolado. Aquelle sentimento identificado pelos experimentos que avaliaram os efeitos de polarização no pequeno grupo (pertencimento, *peer pressure*, viés de confirmação e a sensação de que “todos pensam como eu”), abordado na seção anterior, são reiterados e amplificados nas bolhas digitais, desencadeando no indivíduo a emoção típica da massa. Esse contágio emocional acarreta a correlata regressão da consciência de que fala Freud.

O fenômeno foi descrito com precisão por um dos maiores filósofos da atualidade, Peter Sloterdijk. Para ele, o grande tema da modernidade está na formação de uma subjetividade coletiva racional (o Espírito de Hegel, o proletariado revolucionário de Marx ou ainda a esfera pública emancipada de Habermas). Essa promessa não se realiza historicamente porque em seu lugar entra a massa irracional. O paradoxo está no incremento do individualismo simultaneamente à produção do efeito de massa sobre os indivíduos. É o que Sloterdijk designa como individualismo de massa (“*Massenindividualismus*”) (2000, p. 17). Os meios digitais de comunicação são indispensáveis exatamente na produção desse efeito de massa sem a massa física:

As massas da atualidade deixaram de ser, em essência, massas de aglomeração ou de multidão; elas passaram a um novo regime, em que o caráter de massa se expressa não mais em um concílio físico, mas na participação em programas de mídia de massas. [...] Da massa de aglomeração surge uma massa relacionada a programas – e esta última se emancipou da aglomeração física. Nela, o indivíduo existe como massa. Hoje, faz-se parte de uma massa sem enxergar os outros. (Sloterdijk, 2000, p. 16/17 – grifos meus)

Antes que se erga contra nosso argumento a objeção de especulação filosófica ou abstração sociológica (como se a generalização permitida pela filosofia e pela sociologia não fossem justamente uma forma de alcançar maior precisão na descrição da realidade), vale trazer de imediato as evidências empíricas levantadas de forma sistemática pela psicologia, especificamente em relação ao fenômeno da desinformação. A psicologia cibernética tem avançado profundamente na identificação e na mensuração

objetiva dos efeitos das novas tecnologias digitais sobre a personalidade e a estrutura psíquica (Witty; Young, 2016).

Um dos mais interessantes estudos aponta o chamado “efeito de desinibição *online*” (Suler, 2004), em linha com nosso argumento sobre o efeito regressivo da massa sobre a personalidade individual. A sensação de que podemos dizer qualquer coisa *online*, ou pelo menos que podemos dizer coisas *online* que não diríamos em uma interação cara a cara, é explicada por esse efeito desinibidor desencadeado pela tecnologia – pela unilateralidade da técnica, para recuperar Luhmann. John Suler distingue o efeito benigno (jovens que conseguem, por exemplo, lidar com decepções amorosas e traumas ao se exporem em redes sociais) do efeito *tóxico* de desinibição *online*. O efeito *tóxico* ocorre com a combinação de alguns fatores: (i) anonimato dissociativo: a sensação de que as interações *online* não podem ser diretamente relacionadas à nossa personalidade, quase como se o eu (ego) *online* se tornasse livre do superego – exatamente como sugere a psicologia de massas; (ii) invisibilidade: o fato de não podermos ser vistos ao digitar um tuíte ou um *post* de Facebook pode sugerir maior liberdade para o conteúdo da mensagem; (iii) assincronia: a interação *online* é rápida, mas não é síncrona, e não ter de lidar com a presença do outro ajuda na desinibição; (iv) introjeção solipsística: sem a presença instantânea do outro e sabendo que muito do que postamos sequer será lido, a interação *online* pode simular para o usuário que ele está “falando sozinho”, favorecendo a desinibição; (v) relativização da autoridade: a ausência de sinais externos de status, classe social e autoridade (vestimentas, idade, tom de voz, postura corporal etc.) sugere um nivelamento das interações por uma regra geral de informalidade (Suler, 2004). Como é possível perceber, esse efeito de desinibição *online* é quase a tradução empírica do efeito de massa provocado pela interação *online*: supressão do superego, invisibilidade no anonimato (da massa *online*) etc. O indivíduo é tragado pela massa justamente ao se desinibir dos freios solidificados no superego, deixando-se levar pela emoção coletiva.

A desinibição *online* nada mais é do que uma conduta *emocional* (não racional, pré-cognitiva) para com a tecnologia. Em estudo empírico que analisou respostas cognitivas e afetivas à desinformação política no Facebook, foi possível identificar dois tipos sistemáticos de análise da informação: o escrutínio analítico e a validação heurística. O primeiro é baseado no exame do conteúdo semântico da informação e requer, por isso, tempo e atenção dos usuários; o segundo se guia por regras simples que minimizam o custo de processar a informação (e.g., “consenso = correto”, “informação

nova = correto”, “informação na minha bolha = correto”) (Barfar, 2019, p. 175). A partir disso, o estudo analisou *posts* de Facebook de dez influenciadores conhecidos por difundirem desinformação política (Breitbart, RedState etc.) e de dez veículos jornalísticos centenários (*The New York Times*, *The Wall Street Journal* etc.), mapeando os comentários com um *software* específico (Linguistic Inquiry and Word Count 2015). Os resultados mostram que notícias verdadeiras são processadas sem uma carga afetiva elevada. A desinformação, ao contrário, desencadeia uma alta intensidade emocional. Mas o principal achado é o de que os usuários utilizam o partidarismo como chave heurística para interpretar a informação:

Porque seguidores de uma página de desinformação política (e.g., Breitbart ou Occupy Democrats) no Facebook já se expuseram à desinformação que reforça suas crenças prévias, é menos provável que eles rompam com essa exposição seletiva por meio do escrutínio do mérito e da veracidade da desinformação. [...] Mais precisamente, seguidores de desinformação não são suficientemente motivados para exercer um esforço cognitivo para examinar sua validade [...] Guiados em parte pelo menor esforço no processamento da informação, seguidores no Facebook podem considerar indicações de partidarismo para formar um julgamento em favor da verdade da desinformação política. (Barfar, 2019, p. 175)

Ou seja: dentro de uma bolha em que predomina uma determinada orientação política, os usuários das redes sociais são contaminados pela intensa emoção que circula junto com a desinformação e, *nessa medida*, desativam o processamento cognitivo do conteúdo semântico da mensagem e passam a validá-la em bases emocionais. O que circula nas redes sociais é emoção, e não informação.

Um outro experimento bastante citado pela literatura mostra que a velocidade de replicação de postagens no Twitter sobre temas polarizados de opinião (no caso: controle de armas, união homoafetiva e mudança climática) mostra que o uso de palavras com carga moral ou emocional aumenta a transmissão a uma taxa de 20% por palavra (Brady *et al.*, 2017, p. 7316). O estudo identificou ainda que a difusão é maior em grupos circunscritos do que fora deles – *i.e.*, a mensagem se multiplica rapidamente dentro na bolha, mas não a ultrapassa: “À medida que a difusão de mensagens *online* carregadas de conteúdo emocional-moral é circunscrita por fronteiras de grupo, interações desse tipo parecerão mais provavelmente câmaras de ressonância” (p. 7317).

Ou seja, guiados não pela apreensão cognitiva do conteúdo da mensagem, mas pela emoção que ela desencadeia, os usuários se sentem confortáveis o suficiente em suas bolhas para desativar os freios do superego (efeito de desinibição *online*) e *tomar uma posição no mundo bipolar que a mensagem lhes oferece*. Como fenômeno *social*, a desinformação “ajuda” o usuário a se posicionar politicamente. Por isso, ela precisa que esse ciclo seja constantemente mantido em movimento: “Quem quiser influir sobre ela [a massa], não necessita medir logicamente os argumentos; deve pintar as imagens mais fortes, exagerar e sempre repetir a mesma coisa” (Freud, [1921] 2011, p. 27). É justamente a repetição sistemática que cria o ilusório efeito-verdade da desinformação.

4 REPETIÇÃO, EFEITO-VERDADE, TOMADA DE POSIÇÃO

Finalizando nosso argumento, veremos por que a repetição é importante para o funcionamento social da desinformação. Há uma explicação sociológica para a afirmação (que já se tornou um truísmo), segundo a qual “uma mentira repetida inúmeras vezes se torna verdade”. A repetição é crucial no funcionamento da desinformação (Wardle; Derekhshan, 2017, p. 39). Para desvendar por que, é preciso conjugar a “lógica do inconsciente” com a “teoria do performativo” (Derrida, 1996, p. 21). Aqui utilizaremos a teoria dos atos de fala de Austin, tal como incorporada na teoria da comunicação de Luhmann, para identificar o enunciado performativo contido na mensagem de desinformação política. É a identificação desse enunciado que permite ao usuário posicionar-se em um mundo político bipolar. A desinformação funciona para ele como um redutor de complexidade que lhe permite atribuir sentido e interpretar um mundo complexo. E o destinatário da desinformação aprende a decodificar o enunciado performativo pela repetição da mensagem.

A *força performativa* da linguagem foi identificada de forma canônica por John L. Austin na obra *How to do things with words* (1962), um dos responsáveis pelo giro linguístico da filosofia do século XX. Até então, predominava um entendimento relativamente neutro acerca da linguagem: a linguagem (falada ou escrita) apenas reportaria o mundo, descrevendo-o. Austin percebe que falar é também *fazer* alguma coisa. Austin diferencia atos de fala constativos (“*constative speech acts*”), que simplesmente descrevem um estado de coisas objetivo no mundo (“a terra é redonda”), dos atos de fala performativos (“*performative speech acts*”): aqueles que executam uma ação concreta e transformam o mundo. Exemplos clássicos de

Austin: o “sim” pronunciado na cerimônia de casamento, o testamento, o veredito de um júri e o ato de nomear um navio – há um antes e um depois da palavra pronunciada¹².

Luhmann incorpora os atos de fala de Austin em sua teoria cibernética da comunicação (Luhmann, 1984, p. 197 e ss.). A comunicação, para Luhmann, é composta por três elementos: (i) a *informação*: o conteúdo proposicional, constativo de uma determinada comunicação, *i.e.*, aquilo que ela reporta do mundo fático; (ii) o *enunciado performativo* (“*Mitteilung*”): aquilo que a mensagem opera socialmente, sua intervenção no mundo; e (iii) a *compreensão*: a recepção do conteúdo, isto é, a forma como a mensagem deve ser entendida pelo destinatário. Há, no enunciado performativo, sempre, um apelo para que a mensagem seja entendida de uma determinada maneira. Mas esse apelo é contingente e não está assegurado (do contrário, não existiriam mal-entendidos).

Como essa estrutura conceitual nos ajuda a entender a desinformação?

Vamos partir de um caso célebre de *fake news* no Brasil, que ficou conhecido (infamemente) como “kit gay”. Em 28 de agosto de 2018, o então candidato Bolsonaro apresentou, no Jornal Nacional, o livro *Aparelho sexual e cia.*, da autora francesa Hélène Bruller, sustentando tratar-se de livro distribuído pelo Ministério da Educação para interferir na sexualidade de crianças. O Ministério da Educação chegou a ensaiar política voltada a discutir homofobia nas escolas, mas o programa “Escola sem Homofobia”, idealizado em 2004, foi cancelado pela Presidente Dilma em 2011, ano previsto para seu lançamento. O livro sequer constava do material do programa. Na entrevista, Bolsonaro menciona um suposto “seminário LGBT infantil”, tenta mostrar as páginas do livro dizendo que “estava defendendo as crianças em sala de aula”, porque “um pai não quer chegar em casa e encontrar o filho brincando com boneca por influência da escola”. Menciona *live* no Facebook sobre o tema com 40 milhões de acessos em 15 dias e afirma: “Nada eu tenho contra o gay, eu tenho contra o material escolar em sala de aula”. Por fim, citando o Plano Nacional de Promoção e Cidadania e Direitos Humanos LGBT e a estratégia de desconstrução da heteronormatividade, afirma: “Ou seja, estão ensinando em algumas escolas que homem

12 Há uma intensa discussão acerca da teoria dos atos de fala de Austin que não pode ser recuperada aqui (cf. Bachur, 2017, cap. 1). Para o que segue, basta o uso elementar das três funções dos atos de fala, tal como incorporada por Luhmann.

e mulher tá errado, pode ser sim homem com homem, mulher com mulher, o que é difícil para criancinhas a partir de seis anos de idade”¹³.



Jair Bolsonaro durante a entrevista no 'Jornal Nacional' REPRODUÇÃO / REDE GLOBO

Se analisarmos a mensagem contida nessa fala com auxílio do aparato conceitual de Luhmann e Austin, podemos decantar, a partir das frases de Bolsonaro, algo como *statements* correspondentes aos elementos básicos dos atos de fala – informação, enunciado performativo e compreensão. Informação: “O Ministério da Educação está distribuindo material escolar impróprio para a educação infantil”. Aí está, por assim dizer, a mentira, a informação falsa. Como se sabe, o Ministério não realizou a ação criticada, e o Tribunal Superior Eleitoral determinou a suspensão do uso da expressão “kit gay” pela campanha de Jair Bolsonaro, por se tratar de desinformação¹⁴. Mas há, na mensagem, uma camada adicional de sentido. A força performativa da mensagem está em criar uma clivagem em torno da sexualização precoce de crianças. “O poder das palavras está ligado às imagens que elas evocam, e são de qualquer modo independentes de seu significado real” (Le Bon, 1895, p. 56). Enunciado performativo: “Meu adversário eleitoral defende uma política de interferência na sexualidade de crianças, enquanto eu defendo as crianças em sala de aula”. O enunciado performativo cria um mundo político bipolar em que um tema complexo, carregado de estigma e preconceito, de difícil apreensão pelo público em geral (combate à homo-

13 Vídeo disponível no *link*: <https://globoplay.globo.com/v/6980200/> (ver entrevista a partir do minuto 17, ou o vídeo integral a partir do minuto 19).

14 Cf.: <https://www.tse.jus.br/imprensa/noticias-tse/2018/Outubro/facebook-e-youtube-tem-48-horas-para-retirar-do-ar-vidEOS-com-inverdades-sobre-livro-de-educacao-sexual> (consultado em julho de 2021).

fobia, desconstrução da matriz heteronormativa etc.) é supersimplificado: “Eles acham que homem com mulher é errado e que o certo é homem com homem e mulher com mulher”. Ao criar essa clivagem, a mensagem insta o destinatário a se posicionar. A carga emocional da mensagem (que discutimos na seção anterior) aparece no enunciado performativo. Ele cria um atalho para a compreensão da mensagem: o destinatário não para o que está fazendo e vai checar se o Ministério da Educação distribuiu, de fato, um livro específico; o destinatário mobiliza um saber intuitivo preexistente com o qual ele concorda (“o certo é homem com mulher”) e interpreta a mensagem com esse viés de confirmação. Quando o destinatário “concorda” com a afirmação pressuposta na mensagem, essa concordância “apaga”, por assim dizer, a informação falsa, substituindo-a por uma convergência moral. “Quando os performativos têm sucesso, produzem uma verdade cujo poder se impõe às vezes para sempre” (Derrida, 1996, p. 21). Com isso, o destinatário se desvia do problema real, mas atribui sentido a uma disputa política e consegue se posicionar; toma partido em uma discussão que, nos próprios termos do tema (no caso: a desconstrução da matriz hétero-normativa), ele não consegue manejar. A desinformação oferece ao destinatário uma chave “sim/não”, com a qual ele acredita conseguir se orientar politicamente. Trata-se de redução de complexidade, para usarmos mais uma expressão de Luhmann. Por fim, a compreensão: “Candidatos de esquerda relativizam a sexualidade de crianças”. A repetição de cadeias de desinformação ajuda a sedimentar essa compreensão. Diversos casos subsequentes de desinformação na campanha presidencial de 2018 reforçavam o *statement*: “Candidatos de esquerda relativizam a sexualidade de crianças” (e.g., o caso igualmente infame de mamadeira de apelo sexual supostamente distribuída para crianças). Tudo se passa como se, nessa mensagem, o emissor provocasse o destinatário, instando-o a se posicionar – algo como um “e aí, de que lado você está?”.

Com isso, as pessoas não precisam acreditar na desinformação. É claro que não se descarta essa possibilidade. Mas, como operação *social*, a desinformação prescinde da crença individual para funcionar. Ela circula socialmente *sem* a crença individual. Esse é o ponto. Na bolha digital, a questão da veracidade da informação veiculada se torna secundária diante da convergência moral-emotiva operada pelo enunciado performativo da mensagem. A pesquisa empírica sobre desinformação já elaborou incontáveis *surveys* tentando descobrir se as pessoas realmente acreditam na informação falsa. Há, inclusive, estudos que sugerem que as *surveys* não são a melhor estratégia metodológica porque não é possível aferir se as respostas à própria

survey são sinceras (Bullock *et al.*, 2015, p. 521). Em um dos experimentos mais citados pela literatura, Adam Berinsky, pesquisador do MIT, identificou três tipos de reações à desinformação: (i) a crença genuína na desinformação (“*genuine beliefs*”); (ii) a indiferença descompromissada que não leva a sério se a informação é verdadeira ou não, tratando tudo como “papo furado” (“*cheap talk*”); e (iii) a reação expressiva (“*expressive responses*”), que é, na verdade, uma tomada de posição. Ou seja, no exemplo citado acima, a desinformação acerca do suposto “kit gay” pode circular porque algumas pessoas realmente acreditam nessa informação, mas também porque parte delas acha simplesmente engraçado e parte delas entende que se trata de um posicionamento político (“em defesa da família”, “anti-PT”, ou contra a quimera da “ideologia de gênero”). As pessoas podem consumir a desinformação de forma instrumental: “Eles enviam um sinal sobre seu apoio (ou falta de apoio) a figuras políticas e políticas públicas” (Berinsky, 2017, p. 211).

Essa tomada de posição expressiva não ocorre em um único momento. Ela depende, como já visto, da fragmentação da esfera pública e da interação na bolha digital. Nesse contexto, a repetição da mensagem é crucial, porque ela sedimenta e retroalimenta a polarização política, bem como infunde nos usuários uma narrativa que sustenta uma compreensão deturpada da realidade (como no caso da esdrúxula associação entre “marxismo” e “sexualização de crianças”). O conceito de *narrativa* foi integrado no vocabulário cotidiano da disputa em torno de desinformação nas redes sociais (Barbosa, 2019). Não temos como aprofundar nessa discussão conceitual agora. O importante é compreender *narrativa* não como uma mera “disputa de versões”, mas como o encadeamento de enunciados performativos que sedimentam uma compreensão articulada da realidade. Por exemplo: no caso do chamado “kit gay”, a discussão quanto à falsa ou efetiva distribuição do livro não configura uma disputa de narrativas. A narrativa, em sentido técnico, é o macroenunciado *latente* que a repetição de peças de desinformação (kit gay, mamadeira com bico fálico etc.) sedimenta na pré-compreensão dos destinatários das mensagens (“Candidatos de esquerda relativizam a sexualidade de crianças”). A narrativa é a camada subterrânea de sentido que a desinformação veicula. Por essa razão, a repetição é importante: ela mobiliza o usuário para um posicionamento político.

Mais uma vez, a pesquisa empírica apresenta sólidos fundamentos sobre o papel da repetição no consumo de desinformação política: a câmara de ressonância repete para o usuário tudo aquilo que ele acredita,

de forma que a familiaridade com determinadas afirmações passa a não ser mais questionada por ele (Unkelbach *et al.*, 2019). Sunstein designa o fenômeno da repetição interna às bolhas com o conceito de “cascata” (“*cascade*”): uma torrente de afirmações relativamente semelhantes, relativamente homogêneas acerca de um determinado tema que polariza opiniões, e que é suportada por nossos pares nas redes sociais, reforçando o viés de confirmação (Sunstein, 2017, p. 98). De fato, a “exposição reiterada a uma mensagem que contém uma posição extrema, com a sugestão de que muitas pessoas pensam dessa forma, provavelmente levará aqueles expostos, e provavelmente predispostos, a acreditar na mensagem” (p. 77). Mas acreditar não é o mais importante: “Na maior parte das vezes, o usuário tende a selecionar e compartilhar conteúdo de acordo com uma narrativa específica, ignorando todo o resto” (p. 124/125). Essa narrativa é construída pela repetição de enunciados performativos encadeados de forma a transmitir uma visão de mundo bipolar, instando o destinatário a se posicionar. A narrativa funciona então, para o usuário, como um mecanismo redutor de complexidade, como um atalho para a compreensão de um mundo complexo. Com base na narrativa, o usuário atribui sentido ao mundo e consegue se posicionar politicamente.

CONCLUSÕES

Este artigo procurou oferecer uma visada interdisciplinar sobre o fenômeno da desinformação. Partimos da hipótese de que a desinformação não deve ser explicada ou compreendida como comportamento individual, mas como uma operação genuinamente social: são as características estruturais da sociedade digital que explicam como e por que a desinformação funciona. Essas características (fragmentação da esfera pública, plataformas digitais baseadas no engajamento e em filtros-bolha, efeito de massa e contágio emocional, repetição de enunciados linguísticos e tomada de posição política em um mundo polarizado) estruturam uma lógica comunicativa própria, objetiva. Isso não significa que “todos acreditam” em *fake news*. A pergunta central parece ser outra. Não devemos nos perguntar “por que as pessoas acreditam em *fake news*”, mas quais as condições sociais que permitem que sequências de informação falsa “ajudem” as pessoas a atribuir sentido ao mundo. Nesses termos, o desafio de combater a desinformação e proteger a democracia adquire seu real relevo.

Esse é o cerne do desafio regulatório, pois a experiência jurídica moderna está associada à punição de comportamentos individuais (de

indivíduos ou organizações, mas sempre um comportamento específico, individualizado). Mas como regular um fenômeno inscrito na estrutura social dos meios eletrônicos de comunicação? Se o Direito cuida de processar cadeias de imputação estabelecendo causalidade e responsabilização (Latour, 2002, p. 294/295), como identificar qualquer imputação causal quando os fluxos de desinformação são produzidos e replicados por milhões de usuários em um estágio de contágio emocional, em um frenesi típico da massa? “Se aceitarmos que o cérebro humano nem sempre trabalha de forma racional, simplesmente disseminar informação de qualidade não é uma resposta” (Wardle; Derkhshan, 2017, p. 28). Como vimos, a imagem idílica de uma internet acêntrica, democrática, livre e isonômica já foi substituída há muito por uma estratificação de caráter algorítmico invisível ao usuário. Alguma regulação é, por isso, indispensável. Propostas regulatórias focadas na punição de comportamentos individuais, à luz de nossas conclusões, podem se desvirtuar em punitivismo e vigilância. Mas a discussão de propostas regulatórias concretas deverá ser adiada para uma outra oportunidade.

Pelo momento, é importante acentuar o caráter social da desinformação. Ela funciona socialmente porque permite que os destinatários das mensagens interpretem o mundo de uma forma compatível com seus próprios valores e suas crenças pessoais pré-existentes. Ela funciona, ainda, ao permitir um posicionamento político em um mundo polarizado: é possível definir de que lado se está. A desinformação, por mais paradoxal que possa parecer, oferece uma visão de mundo ao destinatário da mensagem falsa que é compatível com a sua visão de mundo pessoal. Paradoxalmente, ela torna o mundo coerente:

Por trás do aparente absurdo das *fake news* e das teorias da conspiração, oculta-se *uma lógica bastante sólida*. Do ponto de vista dos líderes populistas, as verdades alternativas não são um simples instrumento de propaganda. Contrariamente às informações verdadeiras, *elas constituem um formidável vetor de coesão*. [...] Na prática, para os adeptos dos populistas, a verdade dos fatos, tomados um a um, não conta. O que é verdadeiro é a mensagem *no seu conjunto*, que *corresponde a seus sentimentos e suas sensações*. (Da Empoli, 2019, p. 16 – grifos acrescentados)

Foi justamente essa “lógica sólida” subjacente ao absurdo das *fake news* que procuramos desvendar neste artigo, ainda que inicialmente. O paradoxo sociológico da desinformação está no fato de que a informação falsa gera mais *coesão social* e mais *integração* do que a informação verda-

deira. A verdade gera, “no máximo”, debate público racional – e, com ele, dissenso. Essa lógica sólida oferece um desafio para a democracia, pois ela compromete o fundamento da vida democrática: a liberdade de expressão e o discurso público livre. E não haverá solução fácil para o problema das *fake news*. Será preciso envolver agentes públicos, plataformas e especialistas em busca de um arcabouço institucional que mitigue os efeitos deletérios da desinformação. A descrição do fenômeno e a identificação mais precisa de seus contornos constituem apenas o primeiro passo nesse desafio.

REFERÊNCIAS

- ADORNO, Theodor W.; HORKHEIMER, Max. (1947). *Dialektik der Aufklärung*. Frankfurt am Main: Fischer, 2003.
- APPLEBAUM, Anne. *O crepúsculo da democracia*. Trad. A. Bonruquer. Rio de Janeiro: Record, 2020.
- ARENDT, Hannah. Lying in politics: reflections on the Pentagon Papers. In: ARENDT, H. *Crises of the Republic*. New York: Harvest, p. 3-47, 1972.
- ATHEY, Susan; MOBIUS, Markus; PAL, Jenó. The impact of aggregators on internet news consumption. *National Bureau of Economic Research Working Paper*, 28746, 2017. 37 p.
- AUSTIN, John L. *How to do things with words*. Oxford: Oxford University Press, 1962.
- BACHUR, J. P. *Schrift und Gesellschaft: Die Kraft der Inskriptionen in der Produktion des Sozialen*. Weilerswist: Velbrück Wissenschaft, 2017.
- BARBOSA, Mariana (Org.). *Pós-verdade e fake news: reflexões sobre a guerra de narrativas*. Rio de Janeiro: Cobogó, 2019.
- BARFAR, Arash. Cognitive and affective responses to political disinformation in Facebook. In: *Computers in Human Behavior*, 101:173-179, 2019.
- BENKLER, Yoshai; FARIS, Robert; ROBERTS, Hal. *Network propaganda: Manipulation, disinformation, and radicalization in American Politics*. Oxford: Oxford University Press, 2018.
- BENKLER, Yoshai; SUNSTEIN, Cass et al. The Science of Fake News. *Science*, 359: 1094-1096, 2018.
- BERINSKY, Adam J. Telling the truth about believing the lies? Evidence for the limited prevalence of expressive survey response. *The Journal of Politics*, 80:211-224, 2017.
- BLACHNIO, Agata; PRZEPIORKA, Aneta; PANTIC, Igor. Association between Facebook Addiction, self-esteem and life satisfaction: a cross-sectional study. In: *Computers in Human Behavior*, 55:701-705, 2016.

- BOURDIEU, Pierre. L'emprise du journalisme. In: *Actes de la recherche en sciences sociales*, 101-102:3-9, 1994.
- BULLOCK, John G.; GERBER, Alan S.; HILL, Seth J.; HUBER, Gregory A. Partisan bias in factual beliefs about politics. *Quarterly Journal of Political Science*, 10: 519-578, 2015.
- BRADY, William J.; WILLS, Julian A.; JOST, John T.; TUCKER, Joshua A.; VAN BAVEL, Jay J. Emotion shapes the diffusion of moralized content in social networks. In: *Proceedings of the National Academy of Sciences of the United States of America*, v. 114, n. 28, p. 7.313-7.318, 2017.
- BRANCO, Sérgio. Pequeno guia para identificar *fake news* na campanha eleitoral 2018 (ou como não ser vítima das notícias falsas). In: *VvAa, Fake news e as eleições de 2018*, Rio de Janeiro: Fundação Konrad Adenauer, p. 157-168, 2018.
- CANETTI, Elias. (1960). *Massa e poder*. Trad. S. Tellaroli. São Paulo: Companhia das Letras, 1995.
- DA EMPOLI, Giuliano. *Os engenheiros do caos*. Trad. A. Bloch. São Paulo: Vestígio, 2019.
- DERRIDA, Jaques. História da mentira: prolegômenos. Trad. Jean Briant. In: *Estudos Avançados*, 10, 27: 7-39, 1996.
- EVANGELISTA, Rafael; BRUNO, Fernanda. WhatsApp and political instability in Brazil: targeted messages and political radicalisation. *Internet Policy Review*, v. 8, n. 4: 1-23, 2019.
- FESTINGER, L.; PEPITONE, A.; NEWCOMB, T. Some consequences of deindividuation in a group. *The Journal of Abnormal and Social Psychology*, 47: 382-389, 1952.
- FREUD, Sigmund. (1921). Psicologia das massas e análise do eu. In: *Obras completas*. 6. reimp. São Paulo: Companhia das Letras, v. 15, 2011.
- HABERMAS, Jürgen. *Strukturwandel der Öffentlichkeit: Untersuchungen zu einer Kategorie der bürgerlichen Gesellschaft*. Frankfurt am Main: Suhrkamp, 1962.
- _____. *Faktizität und Geltung: Beiträge zur Diskurstheorie des Rechts und des demokratischen Rechtsstaats*. 4. ed. Frankfurt am Main: Suhrkamp, 1992.
- HATFIELD, Elaine; CACIOPPO, John T.; RAPSON, Richard L. *Emotional contagion*. Cambridge: Cambridge University Press, 1994.
- ICO (INFORMATION COMMISSIONER'S OFFICE). *Democracy disrupted? Personal Information and Political Influence*. London, 2018.
- JACK, Caroline. *Lexicon of lies: terms for problematic information*. Data & Society Research Institute, p. 22, 2017.

- KOSELLECK, Reinhart. *Kritik und Krise: Eine Studie zur Pathogenese der bürgerlichen Welt*. 8. ed. Frankfurt am Main: Suhrkamp, 1959.
- LAPIDOT-LEFLER, Noam; BARAK, Azy. Effects of anonymity, invisibility, and lack of eye-contact on toxic online disinhibition. In: *Computers in Human Behavior*, 28: 434-443, 2012.
- LATOURETTE, Bruno. *La fabrique du droit*. Paris: La Découverte, 2002.
- LE BON, Gustave. *Psychologie des foules*. Paris: Presses Universitaires de France, 1895.
- LUHMANN, Niklas. *Soziale systeme*. Frankfurt am Main: Suhrkamp, 1984.
- _____. *Die Kunst der Gesellschaft*. Frankfurt am Main: Suhrkamp, 1995a.
- _____. (2004). *Die Realität der Massenmedien*. 3. ed. Wiesbaden: VS, 1995b.
- _____. *Die Gesellschaft der Gesellschaft*. Frankfurt am Main: Suhrkamp, 1997.
- NOBRE, Marcos. *Choque de democracia: razões da revolta*. São Paulo: Companhia das Letras, 2013 (e-book).
- O'NEIL, Cathy. *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Broadway Books, 2016.
- PARISER, Eli. *The Filter-Bubble*. New York: Penguin Books, 2011.
- SCHUMPETER, Joseph. *Capitalismo, socialismo e democracia*. Trad. Ruy Jungmann. Rio de Janeiro: Fundo de Cultura, 1961.
- SIMMEL, Georg. (2016). Zur Psychologie und Soziologie der Lüge. In: SIMMEL, G. *Individualismus der modernen Zeit*. 2. ed. Berlin: Suhrkamp, p. 83-94, 1899.
- SLOTERDIJK, Peter. *Die Verachtung der Massen: Versuch über die Kulturkämpfe in der modernen Gesellschaft*. Frankfurt am Main: Suhrkamp, 2000.
- SRNICEK, Nick. *Platform capitalism*. Cambridge: Polity Press, 2017.
- SULER, John. The online disinhibition effect. In: *Cyberpsychology & Behavior*, v. 7, n. 3, p. 321-326, 2004.
- SUMPTER, David. *Outnumbered: From Facebook and Google to fake news and filter-bubbles – The algorithms that control our lives*. London: Bloomsbury Sigma, 2018.
- SUNSTEIN, Cass. *Going to extremes: how like minds unite and divide*. Oxford: Oxford University Press, 2009.
- _____. *#republic: divided democracy in the age of social media*. Princeton: Princeton University Press, 2017.
- TALISSE, Robert. *Overdoing democracy*. Oxford: Oxford University Press, 2019.

TRÄSEL, Marcelo. A eficácia da checagem de fatos no combate à desinformação. In: *VvAa, Fake News e as eleições de 2018*, Rio de Janeiro: Fundação Konrad Adenauer, p. 69-87, 2018.

TUCKER, Joshua A. et al. *Social media, political polarization, and political disinformation: a review of the scientific literature*. Report Hewlett Foundation, 2018. 95 p.

UNKELBACH, Christian; KOCH, Alex; SILVA, Rita R.; GARCIA-MARQUES, Teresa. Truth by repetition: explanations and implications. In: *Current Directinos in Psychological Science*, 28(3): 247-253, 2019.

VESTING, Thomas (2011 a, 2011b, 2013, 2015). *Die Medien des Rechts*, Weilerswist, Velbrück Wissenschaft, v. 1 (*Sprache*), v. 2 (*Schrift*), v. 3 (*Buchdruck*) e v. 4 (*Computernetzwerk*), respectivamente.

_____. A mudança da esfera pública pela inteligência artificial. In: ABBOUD, G.; NERY JR., G., N.; CAMPOS, R. (Org.). *Fake news e regulação*. São Paulo: RT, p. 91-108, 2018.

VILMER, J-B Jeangène; ESCORCIA, A.; GUILLAUME, M.; HERRERA, J. *Information manipulation: a challenge for our democracies*. Report Policy Planning Staff (CAPS), Ministry for Europe and Foreign Affairs & Institute for Strategic Research (IRSEM). Paris, 2018.

VOSOUGHI, Soroush; ROY, Deb; ARAL, Sinan. The spread of true and false news online. *Science*, 359: 1146-1151, 2018.

WARDLE, Claire; DERAKHSHAN, Hossein. *Information disorder: toward an interdisciplinary framework for research and policy making*. Council of Europe Report DGI(2017)09. Strasbourg, 2017. 109 p.

WITTY, Monica; YOUNG, Garry. *Cyberpsychology: the study of individuals, society and digital technologies*. Oxford: Blackwell, 2016.

Sobre o autor

João Paulo Bachur | *E-mail:* joao.bachur@idp.edu.br

Graduado em Direito pela Universidade de São Paulo – USP (2001). Formou-se profissionalmente como Advogado no Escritório Lilla, Huck, Otranto, Ribeiro, Camargo & Messina, em São Paulo, tendo atuado nas áreas de Direito Societário e Tributário (2000-2002). Mestrado (2004) e Doutorado (2009) em Ciência Política pela Faculdade de Filosofia, Letras e Ciências Humanas da USP (2009). Foi Pesquisador Visitante do Instituto de Filosofia da Universidade Livre de Berlim com bolsa de pós-doutorado da Fundação Alexander von Humboldt (2012-2013). Atuou como Professor Voluntário do Instituto de Ciência Política da Universidade de Brasília – UnB (2005-2007, 2010, 2015). Como Servidor Público Federal da Carreira de Especialista em Políticas Públicas e Gestão Governamental (2007-2017), atuou nas áreas de Assessoria Jurídica, Processo Legislativo e Gestão Pública. Foi Chefe de Gabinete do Ministro da Educação (2008-2011), Diretor de Política Regulatória da Secretaria de Regulação e Supervisão da Educação Superior do Ministério da Educação (2015) e Subchefe Adjunto do Núcleo Social da Subchefia de Análise e Acompanhamento de Políticas Governamentais da Casa Civil da Presidência da República (2016). Atualmente, é Advogado e Coordenador do Mestrado e do Doutorado em Direito Constitucional do IDP, em Brasília, e Professor do Insper/SP. Com experiência acadêmica em Direito, Teoria Política, Sociologia e Filosofia da Linguagem, tem publicações nacionais e internacionais nessas áreas.

Data de submissão: 30 de junho de 2021.

Data de aceite: 6 de outubro de 2021.