January 2020 • Working Paper

DEMAND FOR DECEIT

How the Way We Think Drives Disinformation

Samuel Woolley and Katie Joseff

National Endowment for Democracy
Supporting freedom around the world

forum
International Forum for Democratic Studies

Table of Contents

EXECUTIVE SUMMARY

With the advent of digital communications, the rapid spread of political disinformation—the purposeful use of misleading or manipulative information to subvert political discourse and confuse, divide, or otherwise negatively influence publics—has become a vexing challenge for societies worldwide. While political actors have used rumor, incitement, and falsehood as tools since time immemorial, modern technologies allow them to produce and disseminate disinformation at a scale never before seen. The effect has been to blanket the wider information space in confusion and cynicism, fracture democratic societies, and distort public discussion.

Initial assessments of the problem focused on how technological advancements in connectivity have enhanced the reach, speed, and magnitude of disinformation. As a result, the first wave of responses to the challenge was therefore based on the supply of disinformation, and often emphasized fact-checking, debunking, and counter-narratives.

This report focuses on demand for disinformation. While some consumers are exposed to and perhaps influenced by disinformation incidentally, others repeatedly seek out and believe sources of disinformation while rejecting other information sources. Why?

As concerns about computational propaganda and disinformation have rocketed to the top of public agendas, this important question has received greater attention from researchers, policymakers, and civil society. The answer is tied in part to the psychology of news consumption and opinion formation. The lion's share of research has focused on the United States and Europe, but demand-side factors drive the spread and consumption of disinformation around the world. Understanding these factors is crucial for informing effective responses—especially as near future technologies may hold the potential to make these forms of information manipulation even more effective.

Just because the effectiveness of disinformation may be tied to innate aspects of human psychology does not mean that democratic societies are powerless to respond. Rather, civil society, journalists, and other stakeholders invested in the freedom and openness of the global information space should develop innovative adaptations to the contemporary, disinformation-rich information landscape by bearing in mind key insights from the "demand" side of this challenge:

- **Passive and Active Demand for Disinformation.** Demand for disinformation can be broadly split into two categories of psychological drivers: passive, or those requiring no conscious reasoning process on the part of

the individual, and active, or those guided by an individual's efforts to reach conclusions through cognitive processes. Across geographic contexts, deeply polarized societies with low trust in the media appear more susceptible to these drivers.

- **Disinformation as a Global Phenomenon.** Young and vulnerable democracies deserve greater sustained attention and research on these topics. Much of the research on disinformation from the fields of psychology and communications has focused on the impact in developed democracies. Disinformation is straining societies from Australia to Zimbabwe. More work is needed that accounts for this global context.

- **Accounting for Psychology in Fact-Checking Initiatives.** Fact-checkers face challenges in confronting demand for disinformation: news consumers who are invested in a particular political narrative may be more likely to reject corrective information and rationalize their preexisting beliefs. Continuing research aims to understand this challenge and help fact-checkers better communicate with difficult-to-persuade audiences.

- **Mistrust vs. Media Literacy.** Efforts to improve media literacy are similarly challenged, as news consumers who are heavily invested in false political narratives are often quite knowledgeable about (and skeptical toward) independent media. That said, media literacy programs are not all equal: the most effective take into account the demand-side drivers of disinformation.

- **The Impact of Emerging Technologies on the Disinformation Crisis.** Emerging technologies, including synthetic media, virtual and augmented reality, and biometric-powered mass surveillance have the potential to worsen the disinformation crisis in a number of ways. However, it is not only the sophistication of these technologies that poses the greatest challenge, but the interaction with the demand-side drivers discussed here.

Although democratic societies may have underestimated the complexity that demand-side drivers pose, it is important not to panic. This challenge can be met, but doing so will require more research on behavioral change as it relates to digital disinformation, and more work illuminating why people spread novel forms of disinformation. Undoubtedly, the "supply" side of disinformation continues to warrant both empirical research and investigative journalism. Curbing the worst effects of disinformation will also require a better understanding of demand.

INTRODUCTION

Digital disinformation, until recently a marginal field of study, has quickly become a discipline of major concern for researchers and decision makers in academia, the intelligence community, the military, politics, the private sector, and beyond. State-sponsored spin campaigns, the political use of false news reports, and targeted advertising have coalesced into coordinated online and offline projects aimed at manipulating public opinion. These machinations—referred to as computational propaganda, information operations, or influence operations, among many other names—have histories that can be traced through many of the major political and public events of the last decade. Their use spans government type and geographic location, with researchers reporting in late 2019 that there is now clear evidence of such campaigns in more than 70 countries around the world.[1]

A primary intention behind computational propaganda and disinformation is to alter the targeted individuals' perception.

The organized deployment of disinformation tactics on the internet burst into the public consciousness in 2016, with now widely recognized campaigns unfolding in Brazil, the Philippines, South Korea, Syria, Turkey, the United States, the United Kingdom, and elsewhere. Three years later, digital manipulation tools and practices are essentially ubiquitous. Disinformation has become a deeply troubling but deeply ingrained part of the fabric of modern communication.

The global focus on digital disinformation—particularly the exploitation of incendiary rumors, algorithms, automation, and big data to manipulate individuals online—continues to crescendo. Revelations of recent online disinformation campaigns during elections, natural disasters, and security crises in countries from Australia to Zimbabwe have forced policymakers, journalists, software engineers, and citizens alike to pay serious attention to the maleficent potential of social media and other emerging technologies.

Beyond elections, disinformation has fueled ethnic violence, terrorist attacks, and genocide. Powerful political actors—candidates, corporations, militaries, partisan media, incumbent regimes, and special interest groups—as well as collectives of digitally savvy citizens have made effective use of similar tactics to artificially and covertly amplify their messages while attacking their opponents.[2] Both democratically elected and authoritarian governments now engage in digital trolling of their own citizens.[3] Entities ranging from the Canadian House of Commons to the human rights watchdog Freedom House have argued that the very foundations of democracy—including free elections, freedom of expression, and freedom of the press—are under threat.[4]

A primary intention behind computational propaganda and disinformation is to alter the targeted individuals' perception. Recent studies have shown that strategies of sowing false information can also lead to changes in audiences' behaviors.[5] Such efforts do not exist in a vacuum: they influence and are influenced by the supply of and demand for

DEFINING DISINFORMATION

Disinformation, *often used interchangeably with propaganda, is a broad term usually referring to the purposeful use of non-rational argument to undermine a political ideal, inflame social division, or engender political cynicism. It may contain a blend of truth and falsehood, or purposefully exclude important context. Propaganda tends to refer to the use of non-rational argument to either undermine a political ideal or promote a preferred alternative.*

This report deals primarily with ***digital disinformation,*** *or disinformation spread using modern information communications networks.*

Misinformation *refers to the incidental, accidental spread of untrue or misleading information.*

Computational propaganda *refers to the use of computer software to spread and amplify disinformation and otherwise distort or manipulate public conversations through similar tactics, often relying on automation to produce and disseminate content at large scales.*

information. Given that technological advancements in connectivity have enhanced the reach, speed, volume, and arguably the persuasiveness of propaganda, much has been written about the supply side. This report focuses on demand.

The general demand for disinformation is tied to the psychology of information consumption and opinion formation. Especially relevant are the core issues and theories associated with cognitive bias, such as attitude polarization, confirmation bias, and illusory correlation.[6] These concepts address the question of why users seek out and believe some sources of information, whether online or offline, while rejecting others. As concerns about digital propaganda and disinformation have rocketed to the top of public agendas, this important question has received renewed attention from researchers, policymakers, and civil society. The lion's share of this scrutiny has centered on the United States and Europe, but the problem is a global one—demand-side factors drive the spread and consumption of disinformation around the world. Understanding these factors is crucial for developing informed and effective responses.

With this goal in mind, the present report is organized into four parts. Part I discusses the demand-side factors that drive consumption of disinformation and the changes in perception that can result. These factors are separated into "passive" and "active" biases and effects. Part II examines the influence of social, political, and cultural contexts upon this demand. It features two country examples, Mexico and North Macedonia, and describes the current literature on region-specific consumption of disinformation and current counter-disinformation activities. Part III addresses the psychological research concerning corrective measures, focusing in particular on fact-checking and media literacy, which are relevant to the country examples. Part IV briefly explores how near-term technological developments might interact with demand-side drivers of disinformation. The report concludes with recommendations for future research.

PART I: DRIVERS OF DEMAND FOR DISINFORMATION

Demand-side factors drive the spread and consumption of disinformation around the world. Understanding these factors is crucial for developing informed and effective responses.

When people interact with disinformation, several biases inform their desire to consume, share, and internalize its content, as well as their ability to evaluate its veracity. These biases are not categorically faulty: the consumption and evaluation of information is taxing, and human biases have emerged throughout our evolution to enhance decision making, the formation of social bonds, and group cohesion, among other outcomes.[7] For instance, truth bias—the default assumption that information is credible—helps create social trust, enabling efficient communication and economic and societal cooperation.[8] However, it also increases vulnerability to manipulation through disinformation, particularly when one is experiencing a high cognitive load (the number of thoughts that must be held in one's short-term memory simultaneously) or time constraints, since active cognitive engagement is often required to identify and reject false or misleading information.[9]

The psychological biases and heuristics involved in disinformation consumption and perception can be divided into two categories: "passive," meaning ostensibly subconscious reactions, and "active," meaning those that occur during conscious processing of information. While this distinction is not always precise, it enables us to think critically about the efficacy of various countermeasures, such as media literacy and the supply of corrective information through debunking or fact-checking.

Passive Drivers

SHARING DISINFORMATION

The consumption and spread of disinformation owes a great deal to "virality" and the desire of individuals to share emotionally provocative information.[10] Media content, such as fake news articles and Twitter posts,[11] are shared more often and more quickly if they arouse emotion. The type of emotional response provoked also affects virality. Information that evokes high-arousal emotions like fear, disgust, awe, and anger is shared more than information that stimulates low-arousal emotions like sadness.[12] Indeed, individuals who share information online may be more motivated by the prospect of eliciting an emotional reaction in others than by the desire to share true information, further enhancing the spread of disinformation.[13]

EVALUATION OF DISINFORMATION

There are several subconscious biases that can mislead an individual's evaluation of disinformation and engender faulty assumptions of accuracy. For example, exposure to subconscious stimuli that an individual is not aware of, but that are still encoded as memories, can "prime" that individual, shaping perceptions and behavior.[14] Such priming can interact with biases like racial prejudice,[15] strengthening faulty assumptions based on limited and false information.[16]

Priming can be used for persuasion when the stimulus aligns with a properly motivated individual's goals.[17] For instance, research has shown that individuals subliminally primed with a sad face were more enticed by advertisements for mood-enhancing music, but only when they expected to interact with another person and were therefore incentivized to improve their mood.[18] Priming has been found to shape beliefs regarding political information.[19]

COGNITIVE DRIVERS OF CONSUMPTION, ACCEPTANCE, AND SHARING OF DISINFORMATION

PASSIVE DRIVERS

Belief Perseverance Effect: *Continued influence of initial conclusions (sometimes based on false, novel information) on decision-making and individual beliefs.*

Familiarity Effect: *Information which is repeated or delivered in a manner consistent with past experience (for example, in a frequently-heard accent) is often deemed more credible.*

Misinformation Effect: *False information suggested to individuals after the fact can influence their perception, especially as time passes and the memory weakens.*

Priming: *Shaping an individual's perceptions and behavior through exposure to subconscious stimuli.*

Repeat Exposure: *Individuals may respond more positively to stimuli that they have seen frequently than to stimuli they have seen only a few times; persists even when exposure is subliminal and individuals are unaware that they have seen a stimulus.*

Truth Bias: *The default assumption that information is credible.*

Virality and Heightened Emotion: *Information which evokes fear, disgust, awe, anger, or anxiety may be much more likely to be spread by individuals over social media.*

ACTIVE DRIVERS

Bandwagon Effect: *The tendency of individuals to be more likely to adopt beliefs that they believe are common among others.*

Confirmation Bias: *Suggests that individuals seek out information that is in agreement with their preexisting beliefs.*

Consensus Bias: *The tendency to believe information that is perceived as consensus.*

Disconfirmation Bias: *Suggests that people actively reason against information which conflicts with preexisting beliefs.*

Directionally Motivated Reasoning: *The desire to reach a specific conclusion, and thus to lend more credibility to information favoring that conclusion.*

In-group favoritism: *The tendency to favor one's "in-group" (e.g. race, gender, sexual orientation, religious preference, partisan affiliation, geographic location, etc.) over one's out-group.*

Preference Falsification: *Occurs when individuals express preferences (e.g. favored politician or policy) in response to perceived societal pressures and do not communicate their true opinion.*

Prior Attitude Effect: *Suggests that people regard information that supports their beliefs ("pro-attitudinal information") as more legitimate than counter-attitudinal information (sometimes called the prior attitude effect).*

Repetition is another factor which can affect subconscious receptivity to disinformation. People respond more positively to stimuli that they have seen frequently than to stimuli that they have seen only a few times.[20] This "mere-exposure effect" persists even when exposure is subliminal and people are unaware that they have seen a stimulus,[21] and it endures across cultures, species, and stimulus types.[22]

Interestingly, the more positively one views a stimulus, the more familiar the stimulus feels, even if one has never seen it before—leading to a related "good-is-familiar effect."[23] This effect has an important influence on credibility: for example, people often judge information to be more credible if it is spoken in a familiar accent as opposed to an unfamiliar accent.[24] When pervasive disinformation is viewed repeatedly,[25] it can become more familiar to audiences and thus more credible to them.[26] Put another way, increased exposure to unbelievable news headlines can make these headlines seem more believable.[27]

This relationship between the familiarity of disinformation and its credibility should be considered carefully by fact-checkers and those designing public information campaigns intended to correct false information.

This relationship between the familiarity of disinformation and its credibility should be considered carefully by fact-checkers and those designing public information campaigns intended to correct false information.[28] In environments where free and accurate media sources are more prominent than outlets of "fake news," the effects of familiarity may favor accurate information over disinformation. The situation is a complex one, however, with recent research in the journal *Science* suggesting that news often spreads more quickly and widely when it is false.[29]

BELIEF IN DISINFORMATION

The tendency to maintain belief in disinformation in spite of contrary information is attributed to a psychological phenomenon known as "belief perseverance." Initial conclusions based on false but novel information can continue to influence decision making and belief even after they are proven to be unsubstantiated.[30] Perhaps counterintuitively, asking people to think critically about these beliefs can strengthen them: belief perseverance has been found to be stronger and more resistant to correction when individuals are asked to explain how the false information could possibly be true.[31]

A variation of this phenomenon is the "backfire effect"—the idea that belief perseverance can become stronger when new information challenges one's deep-seated beliefs. In one example, when conservatives who believed that a tax cut enhances government revenue received evidence to the contrary, they were found to become more steadfast in their belief than conservatives with the same views who did not receive the corrective information.[32]

It should be noted that the backfire effect is still subject to debate.[33] Research into the effects of strong identity markers (like partisan affiliation) and critical thinking on receptivity to corrective information is ongoing; one hypothesis is that the average person's perceptions of news and consumption patterns are due largely to "lazy information processing" as opposed to motivated reasoning, meaning the backfire effect may primarily manifest in a small portion of the population that scores highly on critical reasoning

tests and has strong preexisting views.[34] Other studies have found that partisanship and motivated reasoning can actually be a crucial factor in explaining the spread of "pipe-dream" fake news—or news that fulfills the wishes of the consumer.[35]

Belief in disinformation can also result from altered memories of an original event. For example, witnesses to a car accident have been shown to reconstruct their memories based on subsequent misinformation: those told after the fact that two cars had "smashed" together claimed to remember the presence of broken glass at the scene, even though there was none.[36] This "misinformation effect" has been found to intensify as the time between the event and the eyewitness's exposure to misinformation increases, presumably because the original memory weakens.[37] The credibility of the individual providing the misinformation influences the magnitude of the effect as well.[38] The effect has been demonstrated in people of all ages, with a wide range of eyewitness events, a variety of misinformation topics, and different methods of misinformation sharing (including face-to-face and written communication). It even occurs when the initial memories are of video recordings as opposed to personal experiences.[39]

SELECTIVE EXPOSURE, FILTER BUBBLES, AND ECHO CHAMBERS

*While the bulk of this paper focuses on cognitive drivers of demand for disinformation, technology also plays a role. In particular, social media platforms may encourage **selective exposure**, a process by which individuals are primarily exposed to information aligning with their prior beliefs and do not interact with information that challenges those beliefs.[40]*

*Selective exposure can facilitate the creation of informational **echo chambers**, self-selected information environments created by joining groups or following certain types of news.[41] Algorithmic sorting carried out by social media platforms and other services can enhance this process, for example through their automated recommendations, creating ideological **filter bubbles**.[42]*

Some believe that echo chambers and filter bubbles not only confirm individuals' existing beliefs, but can also exacerbate extremity of belief, and by extension, overall polarization.[43] Because of the passive psychological drivers of disinformation, a more polarized public may in turn be more easily manipulated. Others assert that concerns about echo chambers are overblown because most people do not consume much political news and those who do consume large amounts of political news tend to access a wide variety of sources.[44] In addition, it has been found that online news consumption is significantly less segregated than offline news consumption from face-to-face interactions.[45] That said, as online communications shift from open platforms like Facebook and Twitter to closed-group messaging applications like WhatsApp and WeChat, selective exposure may become more prevalent.[46]

Active Drivers

CONSUMPTION OF DISINFORMATION

When consciously deciding what information to interact with, a person is subject to a host of biases. Two types of motives driving information consumption are directional motivations (the desire to reach a specific conclusion) and accuracy motivations (the desire to reach the most accurate conclusion).[47] While directionally motivated reasoning can be driven by many factors,[48] two of the main factors with regard to political information, and thus disinformation, are partisanship and preexisting opinions.[49]

While directionally motivated reasoning can be driven by many factors, two of the main factors with regard to political information are partisanship and preexisting opinions.

Directionally motivated reasoning is implicated in several other effects, including "in-group favoritism," which describes the favoring of one's own group (for example, one's race, gender, sexual orientation, religious preference, or geographic location) over others.[50] Partisan bias—preference for one's own political party—is another form of directionally motivated reasoning that can lead to different interpretations of "objective" reality, reinforcing differences between members of rival parties and exacerbating polarization.[51]

The impact of partisan bias on consumption of disinformation is subject to debate, and some experts believe that lack of desire to engage in analytical reasoning, not partisan bias, is the primary reason individuals consume belief-affirming (or, "pro-attitudinal") information.[52] Others believe that individuals do often seek out confirmatory information (confirmation bias), actively reason against incongruent information (disconfirmation bias), and regard pro-attitudinal information as more legitimate than counter-attitudinal information. Collectively, these psychological processes lead to the strongest attitude polarization among those with the strongest prior beliefs and the most biased information-processing tactics.[53]

ENDORSEMENT OF DISINFORMATION

Even if individuals believe a piece of disinformation to be false, they may endorse it as true. "Preference falsification" occurs when people obscure their true opinions, for example about a favored politician or policy, due to societal pressures.[54] Preference falsification is often tied to the "bandwagon effect," in which adherence to a belief or fad increases as more individuals adopt it.

This is in turn often associated with "consensus bias," or the tendency to believe information which is perceived to be the wisdom of the crowd.[55] Preference falsification, in conjunction with the bandwagon effect or consensus bias, can allow regimes and policies to remain in place with ostensible popular support, even if a majority of the populace privately does not support them. Eastern European dissidents under communist rule, for example, described themselves as "living a lie."[56]

Open Questions about Demand-Side Drivers

Biases play a role in information processing and can lead us to consume, share, and believe disinformation. Nevertheless, decision-making about what information to consume, share, and believe is complex, and understanding the role of biases, particularly

in online social networks, requires further research.[57] Current literature continues to debate the extent to which biases impact opinion formation on social media, and the extent to which opinions that are shaped online affect behavior offline. With this in mind, we looked to examples from two different countries to assess known disinformation events from the perspective of psychology.

PART II: COUNTRY-BASED EXAMPLES

The two analyses below investigate how the psychological demand-side drivers of disinformation factor into the broader information environments in Mexico and North Macedonia, including efforts by civil society to counteract misinformation and disinformation. These examinations do not empirically measure demand-side drivers, but rather aim to leverage current literature to suggest how certain drivers may have operated in known cases.

Although disinformation is prevalent throughout the world,[58] Mexico and North Macedonia are both vulnerable democracies where trust in mainstream media and government is low, but civil society activists and organizations are working to counteract false information. They differ in several important ways regarding sources of false information and the media through which it is transmitted. In Mexico, false information is predominately created domestically and distributed through social media and messaging platforms, in part due to high rates of internet penetration.[59] In North Macedonia, a substantial amount of false information comes from foreign sources (Russia in particular) and is distributed through a variety of digital and non-digital media channels.[60]

In the Mexico example, we discuss responses to both emergency-related misinformation and political disinformation by tracing the activities of Verificado 19S, a fact-checking entity that arose out of necessity in the aftermath of the 2017 Puebla earthquake, and Verificado 2018, a more complex fact-checking collaboration that evolved from Verificado 19S in order to monitor the 2018 Mexican general elections. In the North Macedonia example, the focal event is the 2018 referendum on an agreement with Greece to rename the country, which was rife with Russian disinformation.

Mexico

BACKGROUND

As of March 2019, Mexico had the tenth largest population of internet users in the world.[61] Of its estimated 88 million users, 82 million access the internet through smartphones.[62] The three largest telecommunications companies offer unlimited data for Facebook and WhatsApp; as a result, 99 percent of social media users in Mexico use Facebook and 93 percent use WhatsApp. Roughly 24 percent of Mexican WhatsApp users are on the application for six or more hours a day.[63] Despite its limited use in the country, Twitter is influential due to its popularity among politicians and journalists.[64] All in all, despite a level of internet penetration that is lower than the global median,[65] Mexicans are extremely active users of social media.

COMPARING COUNTRY EXAMPLES	MEXICO	NORTH MACEDONIA	SOURCE
Freedom House *Freedom of the Press* Score 0 = least free, 100 = most free	64/100	64/100	*Freedom of the Press,* Freedom House, 2017
Freedom House *Freedom in the World* Score 0 = least free, 100 = most free	63/100	59/100	*Freedom in the World,* Freedom House, 2019
Journalists Killed in 2019	11	0	Committee to Protect Journalists
Journalists Killed in 2010-2019	73	0	Committee to Protect Journalists
Trust in Media	58 percent	14 percent	Edelman Trust Barometer **(Mexico);** Konrad Adenauer Stiftung **(North Macedonia)**
Internet Penetration (percent of population, as of 2018)	65.77 percent	79.17 percent	International Telecommunications Union

Many Mexicans also distrust traditional media and the government. Some 80 percent of respondents in a 2017 survey said that they trusted newspapers a "little" or "not at all."[66] In 2018, 52 percent of Mexicans reportedly distrusted the media, and 72 percent distrusted government.[67] These statistics are comparable to those from other countries in Latin America—61 percent of Argentinians, 57 percent of Colombians, and 57 percent of Brazilians distrust the media, while 41 percent of Argentinians, 24 percent of Colombians, and 18 percent of Brazilians express trust in government.[68]

One contributor to this distrust is the economic model of the media in Mexico, which renders news outlets prone to state control. The majority of funding for newspapers, television stations, and radio stations comes from the government. During the first five years of his presidency, which began in 2012, Enrique Peña Nieto's administration spent nearly $2 billion in federal funds on media advertising; at the state and local levels, members of all parties spend hundreds of millions more. Bribery is rampant and so normalized that some reporters are listed as government contractors.[69] The press is also afflicted by violence: attacks against the press increased by 163 percent between 2010 and 2016, with an additional 11 murders occurring in 2017 alone. Many of these crimes include suspected involvement by public officials.[70] One survey of 102 journalists found that 70 percent had been threatened or attacked due to their work, and 96 percent had colleagues who had been attacked.[71]

Ballot boxes at a polling station in Mexico City during the 2018 general elections.

Despite distrust in the media and government, trust in nongovernmental organizations (NGOs) is very high. At 71 percent, Mexico had the highest level of trust in NGOs among 28 countries surveyed by the 2018 Edelman Trust Barometer.[72] This signals that civil society groups such as AJ+ Verifica have the opportunity to be instrumental in holding the government and media accountable through fact-checking and educating the public through media literacy.

It is against this backdrop that two nongovernmental campaigns worked to confront misinformation and disinformation and their demand-side drivers during two major events: the 2017 Puebla earthquake and the 2018 Mexican general elections. We focus on these two events because they illustrate two important facets of the psychology of disinformation and efforts to counteract it. First, the aftermath of the 2017 Puebla earthquake primarily featured nonpartisan misinformation (false information that is not deliberately misleading), while the 2018 Mexican elections were primarily dominated by partisan disinformation (deliberately false information).[73] Second, the two events showcase the evolution of nonprofit efforts to counteract false information—Verificado 19S, the fact-checking response to the earthquake, was the precursor for Verificado 2018, the fact-checking collaboration during the elections.

2017 PUEBLA EARTHQUAKE

On 19 September 2017, Mexico was struck by a magnitude 7.1 earthquake that left 369 people dead, more than six thousand injured, and vast amounts of property damaged—particularly in Mexico City.[74] The earthquake was immediately followed by a deluge of false information.

False rumors, which can be spread either accidentally or purposefully, are extremely common during crises such as earthquakes, as the chaotic conditions often produce "relative collective ignorance and ambiguity" as well as high levels of anxiety.[75] There have been no studies specifically documenting the spread of computational propaganda—meaning all types of online political manipulation—and the effects of emotions like anxiety during the Puebla earthquake. However, an abundance of viral emotional images and stories on social media and messaging platforms appeared to heighten hysteria and increase the unintentional spread of false information. For example, the hashtag #FridaSofia trended on Twitter following reports that a twelve-year-old girl named Frida Sofia had been saved from the rubble of a school, but it was later shown that no such girl existed.[76] In addition, it is possible that truth bias, compounded with collective eagerness to help, stymied rescue efforts and increased anxiety. According to journalist Sandra Barrón Ramírez, flawed reports of a collapsed building on Twitter led so many people to rush and help that they accidentally blocked emergency personnel.[77] There were several other false reports and delayed reposts of emergency calls for help, which led to new waves of panic.

False rumors, which can be spread either accidentally or purposefully, are extremely common during crises such as earthquakes.

Recognizing that lives were at risk due to false information, a group of Mexican journalists created a crowdsourced information project called Verificado 19S to "channel the desire to help to places that need it."[78] More than 250 volunteers, both journalists and civilians, sent in Google forms detailing the locations of gas leaks and structural damage, places to take shelter, and sites for receiving or donating food, water, and clothing. A Twitter account and Google map with live updates was created, and within four days of the project's launch, the page had 4.5 million views.[79] This impressive network of fact-checkers in the field, computer coders updating the map, and journalists disseminating information was assembled on extremely short notice and with very little governmental support.

Barrón Ramírez argues that a key to Verificado 19S's success was the "chain of social knowledge." The "core element of trust" was that the journalists who were consolidating the reports knew, or came to know, the volunteers fact-checking in the field.[80] Having the means to verify information and to present it without sensationalism enabled Verificado 19S to reduce the impact of viral, emotionally manipulative falsehoods. In addition, by creating a network of trustworthy sources who could verify facts on the ground, and then redistributing that information through a centralized outlet, Verificado 19S disrupted possible online echo chambers.

Given that the topic of emergency responses to the earthquake was largely nonpartisan, Verificado 19S did not need to expend much effort to convince the public of the group's objectivity. This is noteworthy in comparison with other fact-checkers, who are frequently obliged to persuade audiences of the accuracy of information that may conflict with their preexisting biases. As discussed below, fact-checking political information

Building damaged by the 2017 Puebla earthquake.

requires not only investigating veracity but also grappling with directionally motivated reasoning (including partisan bias), confirmation bias, and belief perseverance, among other passive and active drivers of mis- and disinformation consumption, promotion, and internalization.

2018 GENERAL ELECTIONS

On 1 July 2018, Mexico held the largest general elections in its history, with over 3,400 positions contested.[81] Although there was concern that Russian disinformation would influence the presidential vote, the majority of disinformation originated within Mexico.[82] In response, a coalition of news outlets and other organizations including Al-Jazeera's AJ+, Animal Político, and Pop-Up Newsroom established a collaborative fact-checking effort called Verificado 2018.[83] The project was in operation for 119 days and worked with eighty partner newsrooms to publish four hundred posts and fifty videos that debunked "fake news."[84] Beyond simple fact-checking, Verificado 2018 had to contend with the many biases that accompany political disinformation, such as partisan bias, confirmation bias, and the virality of sensationalist information.

About 80 percent of the false stories fact-checked by Verificado 2018 were about Andrés Manuel López Obrador, a politician who is often described as a populist and went on to win the presidential vote.[85] One of the most prominent false stories claimed that López Obrador's wife, Beatriz Gutiérrez Müller, was the granddaughter of Heinrich Müller, an infamous senior official in Nazi Germany. There were several stories asserting that López Obrador was being supported by Russia and Venezuela. On YouTube, where a great deal of disinformation circulated, a video from the Russian state television outlet RT was

overlaid with spoofed Spanish subtitles alleging that Russian leader Vladimir Putin had called López Obrador "the next protégé of the regime." The video was actually about a bodybuilder, and it is unknown who created the doctored version.[86] Disinformation also encouraged belief in conspiracy theories: a survey of 1,003 Mexican adults conducted in November 2017 found that about half (53 percent) believed that outgoing president Enrique Peña Nieto and his party had a secret plan to stop López Obrador from becoming president.[87] This is not necessarily surprising, as prior studies have found that high-tension events, such as elections, enhance belief in conspiracy theories.[88]

Despite the appeal of sensationalist disinformation, the public did seek out facts. Over the course of Verificado 2018's 119 days in operation, its website was visited more than five million times, not counting the interactions its fact-checks received when distributed by partner newsrooms; the project's Twitter and Facebook accounts each had more than two hundred thousand followers.[89] Verificado 2018 also operated a WhatsApp hotline where users could send individual requests for fact-checking. In the first two weeks of operation, it received 18,500 messages, 13,800 of which were answered by the four Verificado 2018 staffers on the WhatsApp team.[90] Some ten thousand people subscribed to the Verificado 2018 WhatsApp channel, where they received and shared daily debunks—viral images depicting a news story, with bullet points explaining why the story was true or false.[91] Although Verificado 2018 disbanded after the elections, AJ+ Verifica formed as a result of the project and continues its fact-checking activity.[92]

Extreme partisan bias and polarization create fertile ground for belief in disinformation and make fact-checking more difficult.

Attracting a broad audience and generating trust in political fact-checking is not easy. Extreme partisan bias and polarization create fertile ground for belief in disinformation and make fact-checking more difficult.[93] Today, those who dare to criticize López Obrador, whether they are civil society figures or journalists, are likely to be accused of being a *"fifí"*—a member of a corrupt and self-serving elite.[94] López Obrador has a high approval rating (70 percent as of July 2019) and ardent followers,[95] many of whom are willing to harass critical journalists.[96] For example, on 12 April 2019, when López Obrador stated at a press conference that homicide rates had decreased since he took office, Univision television anchor Jorge Ramos challenged the statement with verified data, responding that if current trends continued, "2019 will be the bloodiest and most violent year in the modern history of Mexico."[97] Although Ramos cited factual data and López Obrador did not, the journalist faced thousands of attacks on social media, coordinated in part through the hashtag #JorgeRamosProvocador.[98]

This refusal to regard fact-based critiques as legitimate is both an outcome and an intensifier of the deep gulf between perceived in-groups and out-groups in Mexico. Meanwhile, corruption—which is a central concern in Mexico, with 91 percent of Mexicans expressing the belief that political parties are either corrupt or extremely corrupt—continues to fester.[99] In addition, because Mexico does not have a strong history of press independence, people assume that the press is biased. As a result, both partisan bias and confirmation bias influence consumption of news, which can lead to selective exposure and intensified echo chambers.

The difficulty of reaching audiences in echo chambers, compounded by the difficulty of overcoming their preexisting biases, limits the power of fact-checking. Even Verificado

2018, which was so well regarded during the elections, reached a relatively small audience—its ten thousand channel subscribers and five million website visits paled in comparison with the 56.6 million Mexicans who voted in the 2018 elections.[100] That said, Verificado 2018's approach is worth considering for future efforts. In particular, the WhatsApp helpline, which provided individuals with an opportunity to directly reach fact-checkers and then share the findings within their own channels, proved particularly effective for cultivating personalized grassroots responses to disinformation, as opposed to simply broadcasting fact-checks to a disengaged audience.[101]

North Macedonia

BACKGROUND

Although the majority of disinformation in North Macedonia, and throughout the Balkans, comes from domestic sources, research suggests that there is a great deal of cross-border disinformation among Balkan countries and that Russia is the leading source of external disinformation in the region. Moscow purposefully stokes ethnic tensions to encourage destabilization; it also promotes nonpolitical disinformation, such as the idea that vaccines are dangerous in North Macedonia, seemingly with the intention of spreading fear and distrust in governmental institutions.[102] The Kremlin exerts influence through state-funded media outlets such as Sputnik, the Russian Orthodox Church, Russian business magnates who are active in the Balkans, and its close ties with the Serbian government and ethnic Serb leaders in Bosnia and Herzegovina.[103]

THE REFERENDUM TO RENAME MACEDONIA

Political events that unfolded in North Macedonia in 2018 serve as a useful window into the demand-side drivers of disinformation in the Balkans. Albania, Montenegro, Croatia, Romania, and Bulgaria were all members of the North Atlantic Treaty Organization (NATO), but the accession bid of Macedonia, as the country was then known, had stalled.[104] Key to its progress toward NATO membership was the resolution of a decades-long dispute with neighboring Greece over the right to the name "Macedonia," which Athens said could imply territorial claims over a Greek region of the same name.[105] The two governments agreed in June 2018 to change the country's name from "Macedonia" to "North Macedonia," and the government in Skopje scheduled a referendum to approve the agreement for 30 September 2018. A campaign to defeat the measure was waged in the open and with strong backing from domestic opposition groups. The #bojkotiram (boycott) camp communicated across multiple media channels, including social media, about its objections to the deal with Greece. In the end, the vote was fraught with disinformation aligned with both domestic sources and the Russian government, which strongly opposed any expansion of NATO.[106]

In the months prior to the referendum, Facebook pages and Twitter accounts emerged to support a public boycott of the referendum, spreading false information about NATO and the West, and provoking interethnic tension in the country, which has a large ethnic Albanian minority.[107] Gruesome images falsely claimed to show women who were beaten by police for opposing the referendum, taking advantage of citizens' emotions to encourage the further spread of rumors and disinformation.[108] Other pages equated

Until 2019, the name Macedonia was used by different parties to refer to a region of northern Greece and the neighboring country now called North Macedonia.

voting in the referendum with participating in fascism and Nazism, for example by sharing an image of German chancellor Angela Merkel with a Hitler-style mustache and the text, "Boycott the genocide of the Macedonian people!" Another post claimed that European and American officials advocating for the name change deal were driven by Russophobia. The posts had several thousand likes and interactions—significant in a country of only two million people.[109]

The claims of Nazism were not merely for the sake of sensationalism—a common narrative stemming from the Balkan wars of the 1990s claims that "fascists" from the NATO alliance are the mortal enemies of the Orthodox Christian Slavic peoples of the region, including the Serbs and the Macedonians. Consequently, allusions to Nazism in North Macedonia stoke ethnic tensions and evoke familiarity through repeated exposure and reference to a historical narrative. The Russian regime has long played a role in promoting this narrative, even going so far as to give an author who has repeatedly equated the West to Nazis an award for "preserving the historical memory of the WWII and for her fight against the falsification of the history and anti-fascist education of the young generations."[110]

One of the boycott hashtags, #Бојкотира (#Boycott), ranked among the top trending hashtags on Twitter in the months prior to the referendum. Over 80 percent of the posts related to the hashtag were retweets, suggesting that it "was heavily amplified

but lacked much original messaging on Twitter."[111] A study conducted by the Transatlantic Commission on Election Integrity found that automated "bot" accounts made up 10 percent of the conversation about the referendum and predominately promoted the boycott.[112] Elsewhere on the internet, forty new profiles advocating for the boycott emerged on Facebook each day, and hundreds of new, short-term websites supporting the boycott were created.[113]

The hashtag was heavily promoted by Macedonian far-right groups, whose highly coordinated efforts included a website allowing users to easily retweet and share specific boycott content. In online echo chambers, the Macedonian boycott advocates found kindred spirits among far-right conspiracy theorists in the United States.[114] Offline, flyers promoting the boycott were handed out on the streets, and campaigners organized rallies with elaborate sound systems. The ultimate outcome was that many voters felt uncomfortable going to vote or discussing how they would like to vote, suggesting that preference falsification was at play.[115]

Although the boycott movement drew ample support from suspected Russian proxies such as the Russian-Greek billionaire Ivan Savvidis,[116] the disinformation that is most easily attributed to Moscow came from Russia's state-funded news outlet Sputnik. In the month prior to the referendum, Sputnik spread disinformation with numerous stories such as "Macedonia on Edge amid Fears of Manipulation in Looming EU, NATO Membership Vote." However, the majority of the stories had very few Facebook interactions,[117] and their effectiveness in shaping the vote is subject to debate.

On referendum day, only 37 percent of voters participated, raising questions about the poll's legitimacy. However, 90 percent of those who cast ballots endorsed the name change and future NATO and EU accession,[118] and the government had deemed the referendum "consultative" rather than "legally binding," meaning a 50 percent turnout threshold did not technically apply.[119] The parliament voted to proceed with the name change, and in February 2019, the country became formally known as North Macedonia.[120]

An unfortunate result of the Balkans' harsh media climate is that investigative reporting and fact-checking, which are often necessary to expose disinformation campaigns, are primarily conducted by independent journalists and nonprofit organizations—such as the Macedonian NGO Truthmeter.mk and the online platform f2n2.mk—that receive funding from U.S. and European government and philanthropic bodies.[121] Research and investigative journalism entities based in the United States or EU countries, including the Organized Crime and Corruption Reporting Project (OCCRP) and the Atlantic Council's Digital Forensics Research Lab (DFRLab), also engage in such work directly.[122] As a result, those seeking to discredit counter-disinformation efforts frequently describe them as Western propaganda.[123]

PART III: IMPLICATIONS FOR CORRECTIVE MEASURES

The Mexican and North Macedonian examples demonstrate the wide range of problems, and the variety of strategies and tactics, associated with computational

propaganda and the rise of disinformation online. Both examples, for instance, reveal the role of emotional and sensational content in encouraging consumption of false information as well as in prompting ordinary users to disseminate it even more widely. A number of different organizations, and a corresponding collection of tools, curriculums, and best practices, have arisen to address these and other challenges stemming from online disinformation. Some are situated within the larger body of work on fact-checking, while others target media literacy.[124] These corrective measures have social and psychological implications. They rely on particular theories of psychological behavior change and have their own strengths and weaknesses. Undoubtedly, more work in these extant areas—in addition to newer and more innovative approaches—is needed to effectively address the enormous problem at hand.

Fact-Checking

There is a growing body of literature that raises questions about the efficacy of traditional fact-checking as a tactic for countering disinformation.

Today, there are at least 188 fact-checking entities in more than 60 countries, and fact-checking is a fast-growing field in Asia and South America.[125] Fact-checking groups have moved beyond merely rating the veracity of politicians' statements, with many also tracking fulfillment of promises and assembling complex databases of verified information and statistics. There is evidence that fact-checking can be effective as a deterrent to mendacity under some circumstances: a 2015 study of U.S. state legislators, for example, found that those who were sent a letter describing the risks to their reputation if they were caught making misleading claims ended up substantially less likely to receive a negative fact-check compared with colleagues who did not receive the letter.[126]

However, there are a number of obstacles and psychological processes at play in interactions with fact-checking. First, to be affected at all, an individual must encounter the fact-check,[127] which is itself a significant hurdle considering the saturated modern media environment and the much-debated influence of filter bubbles.[128] Second, many fact-checkers base their work on the "deficit model," which holds that hostility toward scientific knowledge is due to a lack of understanding, and assume that individuals will change erroneous beliefs upon exposure to corrective information. But these professionals must contend with directionally motivated reasoning and confirmation bias, which may prevent fact-checks from changing minds.[129] There is also debate over the backfire effect—do individuals accept fact-checks due to truth bias, or does directionally motivated reasoning cause them to entrench themselves more deeply in their incorrect beliefs when exposed to contrary information?[130] Recent research suggests that people are happier to encounter fact-checks that support their beliefs, and that even when a fact-check is regarded as legitimate, it may not change political behavior.[131] One study indicates that fact-checks and other metrics of veracity are unlikely to shift a person's voting behavior.[132]

There are ways to make corrective information more effective. Some evidence indicates that such information has greater impact, especially with ideologically entrenched individuals, when it comes from an ideologically aligned partisan.[133] In addition, when individuals are prompted to be "civic minded" and "good citizens," they are less influenced by partisanship in assessing novel information.[134] Nevertheless, there is a growing body of literature that raises questions about the efficacy of traditional fact-checking as a tactic for countering disinformation, with some researchers arguing that it can sometimes cement existing beliefs or that current catalogued efforts have minimal effects.[135]

Media Literacy

Media literacy is based on "active inquiry and critical thinking about the messages we receive and create."[136] Modern efforts to build media literacy often involve five mainstays: "youth participation, teacher training and curricular resources, parental support, policy initiatives, and evidence base construction."[137] While some praise media literacy as a pathway to agency and independent evaluation of false information, others contend that it wrongfully places the onus on the individual, as opposed to social media platforms, policymakers, or civil society experts.[138] That said, it is one of the more popular tools in the counter-disinformation toolbox.

Given how greatly media literacy programs can vary, it is not easy to evaluate their effectiveness.[139] In a meta-analysis of 51 media literacy interventions, they were found to have an overall positive effect and to have a greater impact on knowledge, for example evaluation of the veracity or bias of a media outlet, than on behavior and attitude. In addition, the analysis found that the fewer steps an intervention has, the more effective it may be, probably due to reduced cognitive effort and lack of confusion on the part of the learner. Repeated exposure to an intervention also seems to lead to greater success, perhaps as a positive result of the mere-exposure effect and familiarity.[140]

Confronting disinformation requires more than just increasing the public's technical understanding of the modern media environment.

Not all media literacy programs are created equal: some have no effect, and some may cause harm. Outdated methods like checklists for evaluating websites can lead individuals astray, consume a lot of time, and provide a roadmap for purveyors of disinformation hoping to evade media literacy checks.[141]

Confronting disinformation requires more than just increasing the public's technical understanding of the modern media environment. Consider the fact that political extremists are often highly media literate and able to influence others through search-engine optimization, coordinated mobilization of both human and bot accounts, and manipulation of social media algorithms. An effective strategy for combating disinformation—whether it comes from extremists, antidemocratic governments, or citizen-led groups—may require a combination of diverse and bespoke approaches, including adjustments to history and civics curriculums, changes in the policies of social media platforms, increased governmental regulation of the digital space, and myriad other endeavors.

Technology and digital tools, including an array of applications and plug-ins aimed at fighting disinformation, also have a contribution to make in alleviating the pressure of computational propaganda. There is, though, a serious need for comparative research demonstrating which "fake news" or "bot" detection algorithms are most effective and how such tools can best be used.

Media literacy and fact-checking are simply among the most well-known and studied methods for countering false information, information polarization, and conspiracy theories. These frameworks, the communities associated with them, and their longitudinal corpus of work should therefore play a central role in mitigating the negative effects of disinformation.

PART IV: UNDERSTANDING FUTURE DEMAND-SIDE DISINFORMATION CHALLENGES

The two country examples provided in this paper represent only a fraction of the ways in which demand for disinformation affects politics in the modern world. Technology and social media are swiftly changing. Advancements in artificial intelligence (AI), data storage capabilities, and computer vision, among other fields, are transforming disinformation, leading to alterations in how and where digital propaganda is spread. This section offers a glimpse at the ways in which recent and emerging technological developments may interact with demand-side disinformation challenges.

Generation and Manipulation of Image, Video, and Audio Content

Deepfakes—a portmanteau of "deep learning" and "fake"—are extremely realistic fake videos created by using AI to synthesize facial expressions (including eye gaze, blinking, and mouth movement), head positions, and body movements. Similarly, Deep Voice and other AI voice-manipulation technologies can alter the modulation of voices (conveying emotion, accent, or gender) and generate completely new speech.[142] Deepfake algorithms have been used to create videos ranging from synthetic pornography based on photographs of nonconsenting women to a clip of U.S. president Donald Trump telling Belgium to exit the Paris climate agreement.[143] These technologies are publicly available,[144] and it will soon be impossible to assess the veracity of images, videos, or audio content with the naked eye or ear.[145]

Given their potential to make it far easier to influence audiences' beliefs and behavior, convincing fabrications of this kind could have numerous applications. In the political domain, manipulated and synthesized audio and visual content might be used to affect diplomatic negotiations, incite conflicts, and manipulate elections. In the social domain, they could be employed to exacerbate polarization and demographic divisions or erode trust in institutions, among other outcomes. One can look to lynchings in India provoked by rumors spread on WhatsApp for a sample of the ways in which deepfakes might stoke fears about "outsiders" or minority groups and incite intercommunity violence.[146]

Of primary concern is the arms race between the creators and the detectors of such manipulations.[147] But even if it became possible to consistently detect deepfakes, several issues remain: the population must trust the detection mechanism; the overwhelming volume of material that platforms moderate continues to grow (500 hours of video are uploaded to YouTube every minute, for example);[148] corrective information that comes after exposure may be ineffective; and veracity may have little bearing on how content is received or what influence it has.[149]

Moreover, videos that are merely edited, such as a video that was slowed to make it appear that U.S. House of Representatives speaker Nancy Pelosi had slurred speech, have been deeply divisive.[150] This suggests that the potency of deepfakes owes less to technical complexity than to their manipulation of the psychological drivers of information processing. A recent study assessing individuals' ability to evaluate an image's

authenticity found that confirmation bias significantly affected conclusions, and that typical cues of online credibility—such as the trustworthiness of the original source, the number of likes (bandwagon effect), and the trustworthiness of the source that endorses or shares the image—did not significantly enhance the accuracy of the subjects' assessments.[151]

Big Data and Mass Surveillance

Big data, a term that usually refers to the use of huge datasets to make inferences about the world, is the foundation of modern AI technology. Mass electronic surveillance facilitates the collection of extensive personalized data, including location records, online activity, and the biometrics necessary for facial recognition.[152] Paired with machine learning and other forms of AI, these stockpiles of information enable companies and governments to use predictive analytics on individuals, inferring their likely behaviors.

The same personalized profiles consequently facilitate the manipulation of behaviors. Organizations such as Cambridge Analytica have claimed to carry out "psychological warfare" by targeting voters based on analysis of Facebook data, though there has been little research that demonstrates whether these efforts are effective.[153] As a marketing firm, Cambridge Analytica possessed neither the financial and administrative resources, the capacity for censorship, nor the access to personal data that many governments have. In addition, the company's advertisements could not respond to real-time changes in emotion. In 2015, an early prototype of responsive billboard advertisements appeared on the streets of London, equipped with a Microsoft Kinect camera that could read viewers' emotions and adapt advertisements accordingly.[154] With far more advanced sensors and richer datasets already available today, it is not difficult to imagine a surveillance state in the near future that could gather behavioral response data to create reactive propaganda on social media or in public places.

Even in the absence of censorship and state surveillance, the applications of big data for disinformation are myriad: For behavioral data collection there is web tracking, location tracking, and cross-device tracking. For manipulation there are a wide variety of social media management tools, including programs that use AI to optimize the targeting of advertisements. There are search-engine optimization programs that trick algorithms to alter search rankings. And there are automated chatbot accounts for use in fraudulent grassroots campaigns, or astroturfing.[155] Advertising-informed big data could allow for subliminal priming to make individuals more susceptible to propaganda, and future disinformation could use personal data to even more effectively provoke racial in-group/out-group bias and confirmation bias. When a particular behavioral outcome is not possible to provoke, it is already commonplace for savvy individuals to craft content—a meme or a simple videogame—that is effective enough to dominate online conversations, distracting users from important information. These kinds of disruption and diversion tactics will become even more effective as targeting becomes more refined.

Utilizing AI, it may become possible to determine the most effective propaganda or disinformation for a particular individual and shape the message in real time. Extensive

data profiles paired with natural language processing (a branch of AI focused on allowing computers to understand and replicate human language) can enable social media chatbots to match a target's personality and manipulate that individual through posts and direct messages.[156] Creating and automatically tweaking these bots will become relatively simple, and they may become more effective as the creators collect more data on their performance. Access to this degree of personalized information and the capability to deliver targeted messages could make those seeking to spread disinformation better able to exploit psychological biases as well. Already bots are used to repeat messages with high frequency—which could exploit the mere-exposure, consensus, and bandwagon effects—and to infiltrate and shape the opinions of closed echo chambers.[157] In the worst-case scenario, chatbots could be deployed in efforts to shift social media users' thoughts and perceptions on a massive scale through emotional contagion.[158]

Virtual Reality and Augmented Reality

Increasing adoption of virtual reality (VR), meaning immersive fully simulated experiences, and augmented reality (AR), in which digital content is superimposed on one's view of the real world, will intensify the expansion of media and advertising beyond the screens of computers and smartphones. While consumer-ready VR headsets have not yet become widely adopted, VR is a promising technology in the fields of vocational, athletic, and military training, and for the purpose of virtual workplace meetings and collaboration. The telecommunications firm Verizon is using VR trainings to prepare employees for robberies,[159] professional snowboarders and football players use VR to train their reaction times,[160] and the U.S. Army is creating a massive multiplayer VR conflict simulation to test for different terrains, enemies, and team dynamics.[161] The technology supporting AR is not as well developed as VR, but Apple's mobile operating system is now equipped with ARKit, a programming interface that allows third-party developers to create AR applications for hundreds of millions of Apple devices.[162]

As VR and AR become a common part of everyday life, the information environment will become even more saturated; as a result, information overload, emotional stimulation,

VR AND AR: WHAT ARE THEY?

Virtual Reality: *Simulated experiences which allow users to perceive settings and objects which do not physically exist, and interact with those spaces and objects through computer software. Virtual reality is a quickly-evolving area of technology with applications in many sectors, including entertainment, medicine, and education.*

Augmented Reality: *Experiences in which virtual scenery or objects are combined with the physical world through an interface which provides sensory feedback intended to simulate interaction with those virtual elements.*

Generally speaking, virtual reality ***replaces*** *an individual's perception of the physical world, while augmented reality* ***modifies*** *or* ***adds*** *to that perception.*

A virtual reality headset.

and algorithm-driven engagement are likely to intensify the impact of disinformation campaigns. There are already VR and AR advertisements and experiences. To list a few, Renault made an Oculus VR advertisement that simulated driving a car,[163] the *New York Times* created a free VR documentary about child refugees for Google Cardboard,[164] the AR game Pokémon Go was downloaded over 750 million times in the year after its 2015 release,[165] and NextVR enables fans to enjoy sporting events and concerts from virtual front-row seats.[166] Soon there will likely be external ads placed inside VR and AR experiences. In AR, where the distinction between reality and artifice could hypothetically become imperceptible, the potential for psychological manipulation through advertisements is especially concerning.

Preliminary studies on VR have found that heightened emotions and increased feelings of "presence"—"being there" in the experience—led to enhanced memory encoding.[167] How will one differentiate between memories created in reality and those influenced by ads or other mechanisms in VR and AR? Looking to the misinformation effect, it seems possible that an AR avatar could provide leading prompts after a real event in order to subtly shape an individual's memory and perception. For instance, if an individual witnessed soldiers beating a peaceful protester, the government might intrude on the person's digital tools and cause an AR assistant to provide faulty images of the event after the fact, making it appear that the protester was armed and attacking the soldiers.

Just as chatbots have the potential to provoke mass emotional contagion, so too do VR and AR.[168] Studies of VR's emotional impact have found it to be a viable mechanism for provoking specific emotional responses, such as anxiety through a stressful simulation and relaxation through a peaceful simulation.[169] VR and AR may further enhance the effects of subliminal priming, and avatars in VR are likely to be even more effective than chatbots at nudging belief and behavior shifts through social engagement.[170] VR is already used to create lasting behavior change through exposure therapy.[171] What other types of behavior change could be provoked using these technologies?

CONCLUSION

The psychological literature on the demand-side drivers of disinformation can provide useful insights about potential levers for reducing the impact of such content—whether through established work on fact-checking and media literacy or via new tools and ideas that have yet to be developed. Known examples in which disinformation and misinformation have flowed both online and offline, including the cases detailed here, are helpful in drawing correlations between current disinformation tactics and the relevant strategies and theories from the field of psychology. Significant work needs to be done, however, to truly connect psychological drivers to the spread of disinformation online. Simply put, we need more work that tracks behavioral change as it relates to digital disinformation. We need more research that explains why people spread novel forms of manipulative material. Undoubtedly, the supply side of disinformation continues to warrant both empirical research and investigative journalism. But in order to curb the worst intentions of the perpetrators, we must also develop a better understanding of the demand.

The psychological phenomena described in the country examples, and with regard to future technological advancements, are limited in that they are based on inference from extant details and analogous psychology studies. These forecasts are intended to provoke thought and further research. Because there may be differences across social and cultural contexts, researchers looking to explicitly connect psychological phenomena to disinformation should conduct surveys or experiments specifically crafted to investigate individuals' perceptions of and interactions with disinformation in the regions of interest. Similarly, one of the major problems with extant technologies for tracking disinformation is tied to the fact that these tools are created with one language, platform, or region in mind. Indeed, even the world's largest social media companies have struggled to apply their remedies for election interference or coordinated hate speech in countries with less common languages and more limited pools of accessible sociopolitical expertise. It is in part because of these considerations that research and knowledge from the domains of media literacy and fact-checking—which account for linguistic, geographical, and cultural differences—are particularly useful for those hoping to lead efforts to fight disinformation online.

There is significant space, and pressing need, for future studies on the psychological dynamics of propaganda and persuasion via new technologies, such as a virtual avatar, haptic nudges (like the vibration of a smartwatch), or the targeted delivery of suggestive messages to shape users' perceptions after they view a YouTube video (provoking the misinformation effect). Greater research on the nuances of psychological phenomena, including variation by age, socioeconomic class, country, or region, is also crucial. Researchers must understand how particular groups of people spread misinformation or disinformation, as with a team who found that people over the age of 65 were most active in spreading misinformation on Facebook during the 2016 U.S. election.[172]

Paradoxically, such fine-tuned studies are easier to conduct for actors like governments or large corporations that enjoy greater access to personal data. This is especially true in the most authoritarian settings. Without robust support for research by academics, civil society, and other sectors that are committed to the pursuit of knowledge and the public interest, valuable insights about the demand side of disinformation may be hoarded for the purpose of malign manipulation. For the sake of a healthy public square, it is essential that these insights instead see the light of day.

ENDNOTES

1 Samantha Bradshaw and Philip N. Howard, "The Global Disinformation Order: 2019 Inventory of Organised Social Media Manipulation," working paper, University of Oxford, 2019.

2 Samuel C. Woolley, "Automating Power: Social Bot Interference in Global Politics," *First Monday* 21, no. 4 (2016); Yochai Benkler, Robert Faris, and Hal Roberts, *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics* (Oxford, UK: Oxford University Press, 2018).

3 Carly Nyst and Nick Monaco, *State-sponsored Trolling: How Governments Are Deploying Disinformation as Part of Broader Digital Harassment Campaigns* (Palo Alto, CA: Institute for the Future, 2018); Samantha Bradshaw and Philip N. Howard, "Troops, Trolls and Troublemakers: A Global Inventory of Organized Social Media Manipulation," working paper, University of Oxford, 2017.

4 Bob Zimmer, chair/ed., "Democracy Under Threat: Risks and Solutions in The Era of Disinformation and Data Monopoly," Canadian House of Commons, 24th Parliament, 1st Session, 2018; Freedom House, *Freedom on the Net: The Rise of Digital Authoritarianism* (New York: Freedom House, 2018).

5 Brian G. Southwell, Emily A. Thorson, and Laura Sheble, eds., *Misinformation and Mass Audiences* (Austin, TX: University of Texas Press, 2018).

6 Raymond S. Nickerson, "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises," *Review of General Psychology* 2, no. 2 (1998): 175–220; Charles G. Lord, Lee Ross, and Mark R. Lepper, "Biased Assimilation and Attitude Polarization: The Effects of Prior Theories on Subsequently Considered Evidence," *Journal of Personality and Social Psychology* 37, no. 11 (1979): 2098–2109; Loren J. Chapman, "Illusory Correlation in Observational Report," *Journal of Verbal Learning and Verbal Behavior* 6, no. 1 (1967): 151–155.

7 Dominic D. P. Johnson, Daniel T. Blumstein, James H. Fowler, and Martie G. Haselton. "The Evolution of Error: Error Management, Cognitive Constraints, and Adaptive Decision-making Biases," *Trends in Ecology & Evolution* 28, no. 8 (2013): 474–481.

8 Timothy R. Levine, "Truth-default Theory (TDT) a Theory of Human Deception and Deception Detection," *Journal of Language and Social Psychology* 33, no. 4 (2014): 378–392.

9 Daniel T. Gilbert, Romin W. Tafarodi, and Patrick S. Malone, "You Can't Not Believe Everything You Read," *Journal of Personality and Social Psychology* 65, no. 2 (1993): 221–233.

10 Jonah Berger, "Arousal Increases Social Transmission of Information," *Psychological Science* 22, no. 7 (2011): 891–893.

11 Soroush Vosoughi, Deb Roy, and Sinan Aral, "The Spread of True and False News Online," *Science* 359, no. 6380 (2018): 1146–1151; Stefan Stieglitz, Linh Dang-Xuan, Axel Bruns, and Christoph Neuberger, "Social Media Analytics: An Interdisciplinary Approach and Its Implications for Information Systems," *Business & Information Systems Engineering* 6, no. 2 (2014): 89–96.

12 Angela Dobele, Adam Lindgreen, Michael Beverland, Joëlle Vanhamme, and Robert van Wijk, "Why Pass on Viral Messages? Because They Connect Emotionally," *Business Horizons* 50, no. 4 (2007): 291–304; Jonah Berger and Katherine L. Milkman, "What Makes Online Content Viral?" *Journal of Marketing Research* 49, no. 2 (2012): 192–205.

13 Ellen M. Cotter, "Influence of Emotional Content and Perceived Relevance on Spread of Urban Legends: A Pilot Study," *Psychological Reports* 102, no. 2 (2008): 623–629; cited in Vosoughi, Roy, and Aral, "The Spread of True and False News Online."

14 Ap Dijksterhuis and John A. Bargh. "The Perception-behavior Expressway: Automatic Effects of Social Perception on Social Behavior," *Advances in Experimental Social Psychology* 33 (2001): 1–40.

15 John A. Bargh, Mark Chen, and Lara Burrows, "Automaticity of Social Behavior: Direct Effects of Trait Construct and Stereotype Activation on Action," *Journal of Personality and Social Psychology* 71, no. 2 (1996): 230; Jennifer L. Eberhardt, Phillip Atiba Goff, Valerie J. Purdie, and Paul G. Davies, "Seeing Black: Race, Crime, and Visual Processing," *Journal of Personality and Social Psychology* 87, no. 6 (2004): 876.

16 Mark L. Howe, Emma Threadgold, Jenna Norbury, Sarah R. Garner, and Linden J. Ball, "Priming Children's and Adults' Analogical Problem Solutions with True and False Memories," *Journal of Experimental Child Psychology* 116, no. 1 (2013): 96–103.

17 Erin J. Strahan, Steven J. Spencer, and Mark P. Zanna, "Subliminal Priming and Persuasion: Striking While the Iron Is Hot," *Journal of Experimental Social Psychology* 38, no. 6 (2002): 556–568.

18 Ibid.

19 Rafael Di Tella, Sebastian Galiani, and Ernesto Schargrodsky, "Reality Versus Propaganda in the Formation of Beliefs About Privatization," *Journal of Public Economics* 96, nos. 5–6 (2012): 553–567; Joanne M. Miller and Jon A. Krosnick, "News Media Impact on the Ingredients of Presidential Evaluations: A Program of Research on the Priming Hypothesis," in *Political Persuasion and Attitude Change* (Ann Arbor, MI: University of Michigan Press, 1996), 79–100.

20 Robert B. Zajonc, "Attitudinal Effects of Mere Exposure," *Journal of Personality and Social Psychology* 9, no. 2 pt. 2 (1968): 1–27.

21 S. T. Murphy, J. L. Monahan, and R. B. Zajonc, "Additivity of Nonconscious Affect: Combined Effects of Priming and Exposure," *Journal of Personality and Social Psychology* 69, no. 4 (1995): 589.

22 Robert B. Zajonc, "Mere Exposure: A Gateway to the Subliminal," *Current Directions in Psychological Science* 10, no. 6 (2001): 224–228.

23 Benôit Monin, "The Warm Glow Heuristic: When Liking Leads to Familiarity," *Journal of Personality and Social Psychology* 85, no. 6 (2003): 1035.

24 Shiri Lev-Ari and Boaz Keysar, "Why Don't We Believe Non-Native Speakers? The Influence of Accent on Credibility," *Journal of Experimental Social Psychology* 46, no. 6 (2010): 1093–1096.

25 Philip Ball, "'News' Spreads Faster and More Widely When It's False," *Nature*, March 2018.

26 Ian Maynard Begg, Anne Anas, and Suzanne Farinacci, "Dissociation of Processes in Belief: Source Recollection, Statement Familiarity, and the Illusion of Truth," *Journal of Experimental Psychology* 121, no. 4 (1992): 446–458.

27 Jeffrey A. Gibbons, Angela F. Lukowski, and W. Richard Walker, "Exposure Increases the Believability of Unbelievable News Headlines Via Elaborate Cognitive Processing," *Media Psychology* 7, no. 3 (2005): 273–300.

28 Norbert Schwarz, Lawrence J. Sanna, Ian Skurnik, and Carolyn Yoon, "Metacognitive Experiences and the Intricacies of Setting People Straight: Implications for Debiasing and Public Information Campaigns," *Advances in Experimental Social Psychology* 39 (2007): 127–161; Stephan Lewandowsky, Ullrich K. H. Ecker, Colleen M. Seifert, Norbert Schwarz, and John Cook, "Misinformation and Its Correction: Continued Influence and Successful Debiasing," *Psychological Science in the Public Interest* 13, no. 3 (2012): 106–131.

29 Vosoughi, Roy, and Aral, "The Spread of True and False News Online."

30 L. Ross, M. R. Lepper, and M. Hubbard, "Perseverance in Self-perception and Social Perception: Biased Attributional Processes in the Debriefing Paradigm," *Journal of Personality and Social Psychology* 32, no. 5 (1975): 880–892.

31 Craig A. Anderson, Mark R. Lepper, and Lee Ross, "Perseverance of Social Theories: The Role of Explanation in the Persistence of Discredited Information," *Journal of Personality and Social Psychology* 39, no. 6 (1980): 1037–1049.

32 Brendan Nyhan and Jason Reifler, "When Corrections Fail: The Persistence of Political Misperceptions," *Political Behavior* 32, no. 2 (2010): 303–330.

33 Thomas Wood and Ethan Porter, "The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence," *Political Behavior* 41, no. 1 (2019): 135–163.

34 Gordon Pennycook and David G. Rand, "Lazy, Not Biased: Susceptibility to Partisan Fake News Is Better Explained by Lack of Reasoning Than by Motivated Reasoning," *Cognition* 188 (2019): 39–50.

35 Lauara Farago, Anna Kende, and Peter Kreko, "We Only Believe in News We Doctored Ourselves: The Connection Between Partisanship and Political Fake News," *Social Psychology* (2019).

36 Elizabeth F. Loftus and John C. Palmer, "Reconstruction of Automobile Destruction: An Example of the Interaction Between Language and Memory," *Journal of Verbal Learning and Verbal Behavior* 13, no. 5 (1974): 585–589.

37 Elizabeth F. Loftus, David G. Miller, and Helen J. Burns, "Semantic Integration of Verbal Information into a Visual Memory," *Journal of Experimental Psychology: Human Learning and Memory* 4, no. 1 (1978): 19–31.

38 Vicki L. Smith and Phoebe C. Ellsworth, "The Social Psychology of Eyewitness Accuracy: Misleading Questions and Communicator Expertise," *Journal of Applied Psychology* 72, no. 2 (1978): 294.

39 Maria S. Zaragoza, Robert F. Belli, and Kristie E. Payment, "Misinformation Effects and the Suggestibility of Eyewitness Memory," in *Do Justice and Let the Sky Fall: Elizabeth Loftus and Her Contributions to Science, Law, and Academic Freedom* (Mahwah, N.J.: Erlbaum Associates, 2007), 35–63; Melanie K. T. Takarangi, Sophie Parker, and Maryanne Garry, "Modernising the Misinformation Effect: The Development of a New Stimulus Set," *Applied Cognitive Psychology* 20, no. 5 (2006): 583–590.

40 David O. Sears and Jonathan L. Freedman, "Selective Exposure to Information: A Critical Review," *Public Opinion Quarterly* 31, no. 2 (1967): 194–213.

41 R. Kelly Garrett, "Echo Chambers Online? Politically Motivated Selective Exposure Among Internet News Users," *Journal of Computer-Mediated Communication* 14, no. 2 (2009): 265–285.

42 Eli Pariser, *The Filter Bubble: How the New Personalized Web Is Changing What We Read and How We Think* (New York: Penguin Press, 2011).

43 Seth Flaxman, Sharad Goel, and Justin M. Rao, "Filter Bubbles, Echo Chambers, and Online News Consumption," *Public Opinion Quarterly* 80, (2016): 298–320; Cass Sunstein, *Echo Chambers: Bush v. Gore, Impeachment, and Beyond* (Princeton, NJ: Princeton University Press, 2001).

44 Elizabeth Dubois and Grant Blank, "The Echo Chamber Is Overstated: The Moderating Effect of Political Interest and Diverse Media," *Information, Communication & Society* 21, no. 5 (2018): 729–745; Andrew Guess, Brendan Nyhan, Benjamin Lyons, and Jason Reifler, "Avoiding The Echo Chamber About Echo Chambers," Knight Foundation, 2018.

45 Matthew Gentzkow and Jesse M. Shapiro, "Ideological Segregation Online and Offline," *Quarterly Journal of Economics* 126, no. 4 (2011): 1799–1839.

46 James Andrew Lewis and William A. Carter, "Scoping Law Enforcement's Encrypted Messaging Problem," Technology Policy Blog, Center for Strategic and International Studies, 6 April 2018.

47 Thomas J. Leeper and Rune Slothuus. "Political Parties, Motivated Reasoning, and Public Opinion Formation," *Political Psychology* 35 (2014): 129–156.

48 Ziva Kunda. "The Case for Motivated Reasoning," *Psychological Bulletin* 108, no. 3 (1990): 480.

49 D. J. Flynn, Brendan Nyhan, and Jason Reifler, "The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs About Politics," *Political Psychology* 38 (2017): 127–150.

50 Donald M. Taylor and Janet R. Doria, "Self-serving and Group-serving Bias in Attribution," *Journal of Social Psychology* 113, no. 2 (1981): 201–211.

51 Larry M. Bartels, "Beyond the Running Tally: Partisan Bias in Political Perceptions," *Political Behavior* 24, no. 2 (2002): 117–150; Faragó, Kende, and Krekó, "We Only Believe in News that We Doctored Ourselves," *Social Psychology* (2019); Brian J. Gaines, James H. Kuklinski, Paul J. Quirk, Buddy Peyton, and Jay Verkuilen, "Same Facts, Different Interpretations: Partisan Motivation and Opinion on Iraq," *Journal of Politics* 69, no. 4 (2007): 957–974.

52 Pennycook and Rand, "Lazy, Not Biased."

53 Charles S. Taber and Milton Lodge, "Motivated Skepticism in the Evaluation of Political Beliefs," *American Journal of Political Science* 50, no. 3 (2006): 755–769.

54 Timur Kuran, "Preference Falsification, Policy Continuity and Collective Conservatism," *Economic Journal* 97, no. 387 (1987): 642–665.

55 H. Leibenstein, "Bandwagon, Snob, and Veblen Effects in the Theory of Consumers' Demand," *Quarterly Journal of Economics* 64, no. 2 (1950): 183–207.

56 Timur Kuran, *Private Truths, Public Lies: The Social Consequences of Preference Falsification* (Cambridge, MA: Harvard University Press, 1997).

57 Mehdi Moussaïd, Juliane E. Kämmer, Pantelis P. Analytis, and Hansjörg Neth, "Social Influence and the Collective Dynamics of Opinion Formation," *PLOS ONE* 8, no. 11 (2013): e78433.

58 Samantha Bradshaw and Philip N. Howard, "The Global Disinformation Order: 2019 Inventory of Organised Social Media Manipulation," working paper, University of Oxford, 2019.

59 J. Clement, "Countries with the Highest Number of Internet Users as of March 2019 (in millions)," *Statista*, 2019.

60 Nivedita Arvind, Sara Bak, Alex Buzzell, Noah Durette, Naomi Eguchi Faletti, Sarah Nichols, Sofija Raisys, Omar Tabuni, and Jennifer Yan, *NATO: Building Resiliency and Integrity Against Russian Hybrid Warfare Threats* (Seattle, WA: Henry M. Jackson School of International Studies, University of Washington, 2019).

61 Ibid.

62 Stéphanie Chevalier, "Digital Population in Mexico as of January 2019 (in millions)," *Statista*, 29 July 2019.

63 Vladimir Beciez, telephone interview by Katie Joseff, 10 July 2019; Stéphanie Chevalier, "Reach of leading social networks in Mexico as of May 2019," *Statista*, 10 October 2019; "¿Cómo usan los mexicanos WhatsApp?" Comunicación Política Aplicada, 2019.

64 Vladimir Beciez, telephone interview by Katie Joseff, 10 July 2019.

65 Jacob Poushter, Caldwell Bishop, and Hanyu Chwe, "Social Media Use Continues to Rise in Developing Countries but Plateaus Across Developed Ones," Pew Research Center, Washington, DC, 19 June 2018.

66 "How much do you trust newspapers?" *Statista*, 2018.

67 *2018 Edelman Trust Barometer Global Report* (New York: Edelman, 2018).

68 Ibid.

69 Azam Ahmed, "Using Billions in Government Cash, Mexico Controls News Media," *New York Times*, 25 December 2017.

70 "Mexico and Central America," Article 19, 2019.

71 Jorge Luis Sierra, *Digital and Mobile Security for Mexican Journalists and Bloggers: Results of a Survey of Mexican Journalists and Bloggers* (Washington, DC: Freedom House and International Center for Journalists, 2013).

72 *2018 Edelman Trust Barometer Global Report*.

73 Bernd Carsten Stahl, "On the Difference or Equality of Information, Misinformation, and Disinformation: A Critical Research Perspective," *Informing Science Journal* 9 (2006).

74 "Death Toll in Mexico Earthquake Rises to 369 as Last Body Pulled from Rubble," CBS News, 4 October 2017.

75 Benigno E. Aguirre and Kathleen J. Tierney, "Testing Shibutani's Prediction of Information Seeking Behavior in Rumor," Disaster Research Center, University of Delaware, 2001, cited in Onook Oh, Kyounghee Hazel Kwon, and H. Raghav Rao, "An Exploration of Social Media in Extreme Events: Rumor Theory and Twitter During the Haiti Earthquake 2010," International Conference on Information Systems 2010 Proceedings.

76 Daniel Funke, "After Mexico City's Earthquake, This Site Is Crowdsourcing to Map Emergency Resources," *Poynter*, 23 September 2017.

77 Sandra Barrón Ramírez, telephone interview by Katie Joseff, 12 July 2019.

78 Ana Campoy, "In Both the US and Mexico, Citizens Led Better Disaster Response Than Their Governments," *Quartz*, 30 September 2017.

79 Funke, "After Mexico City's Earthquake, This Site Is Crowdsourcing to Map Emergency Resources."

80 Ibid.

81 Juliana Fregoso, "Mexico's Election and the Fight Against Disinformation," European Journalism Observatory, 27 September 2018.

82 Mia Armstrong, "Mexico's Chapter in the Saga of Election Disinformation," *Slate*, 2 August 2018; Digital Forensic Research Lab (DFRLab), "#ElectionWatch: Russian Bots in Mexico?" *Medium*, 27 May 2018.

83 Kate Linthicum, "Mexico Has Its Own Fake News Crisis. These Journalists Are Fighting Back," *Los Angeles Times*, 15 April 2018.

84 "Hasta Luego, Verificado 2018," Verificado 2018, 9 July 2018.

85 Linthicum, "Mexico Has Its Own Fake News Crisis. These Journalists Are Fighting Back."

86 Fregoso, "Mexico's Election and the Fight Against Disinformation."

87 Jorge Buendia, "Fake Poll as Fake News: The Challenge for Mexico's Elections," Wilson Center, Mexico Institute, April 2018.

88 Jan-Willem van Prooijen and Nils B. Jostmann, "Belief in Conspiracy Theories: The Influence of Uncertainty and Perceived Morality," *European Journal of Social Psychology* 43, no. 1 (2013): 109–115.

89 "Verificado 2018, the News Verification Initiative by AJ+ Español, Animal Político and Pop-Up Newsroom, Wins a World Digital Media Award," *Al-Jazeera*, 24 June 2019.

90 Laura Hazard Owen, "WhatsApp Is a Black Box for Fake News. Verificado 2018 Is Making Real Progress Fixing That," Nieman Lab, 1 June 2018.

91 Ibid.; Verificado 2018.

92 Mark Oprea, "The Spread of Fake News Has Had Deadly Consequences in Mexico. Meet the People Trying to Stop It," *Pacific Standard*, 15 February 2019.

93 Victoria Gaytan, "The High Cost of Polarization for Mexico's Young Democracy," *Global Americans*, 16 May 2019; Vladimir Beciez, telephone interview by Katie Joseff, 10 July 2019.

94 Gaytan, "The High Cost of Polarization for Mexico's Young Democracy."

95 Jack Daniel, "Mexico's López Obrador Ratings Slip: To 70% Approval," Reuters, 17 July 2019.

96 César López Linares, "López Obrador Creates Polarization with Attacks on the Press and Little Transparency, Say Mexican Journalists," Journalism in the Americas, University of Texas Austin, 18 April 2019.

97 Kate Linthicum, "Mexico's López Obrador Says Homicide Rates Are Down, Despite Data to the Contrary," *Los Angeles Times*, 12 April 2019.

98 "López Obrador Creates Polarization."

99 "Corruption and the 2018 Mexico Election: What Comes Next?" Center for the Advancement of Public Integrity, Columbia University, 27 July 2018.

100 Andrea Tanco, "Infographic | 2018 Mexican Presidential Election Results," Wilson Center, Mexico Institute, 10 July 2018, *www.wilsoncenter.org/article/infographic-2018-mexican-presidential-election-results*.

101 "WhatsApp Is a Black Box for Fake News."

102 Tijana Cvjetićanin, Emir Zulejhić, Darko Brkan, Biljana Livančić-Milić, *Disinformation in the Online Sphere: The Case of BiH* (Sarajevo: Citizens' Association "Why Not," April 2019); David Wemer, "The Western Balkans: A Growing Disinformation Battleground," Atlantic Council, 7 March 2019.

103 Paul Stronski, "Is Russia Up to No Good in the Balkans?" Carnegie Endowment for International Peace, 13 February 2019.

104 Janusz Bugajski, "NATO Secures the Western Balkans," Center for European Policy Analysis, 11 February 2019.

105 Helena Smith, "Macedonia Officially Changes Its Name to North Macedonia," *Guardian*, 12 February 2019.

106 Asya Metodieva, "How Disinformation Harmed the Referendum in Macedonia," German Marshall Fund, 2 October 2018.

107 Ibid.

108 Milena Veselinovic, "Macedonia Sees Low Turnout in Name Change Referendum Amid Disinformation Campaign," CNN, 30 September 2018.

109 Vladimir Petreski. "#ElectionWatch: Fascist Falsification Ahead of Macedonian Referendum," DFRLab, Atlantic Council, 28 September 2018.

110 Ibid.

111 Kanishk Karan, "#ElectionWatch: Boycott Campaign in Macedonia Features Familiar Characters," DFRLab, Atlantic Council, 14 September 2018.

112 Christina Maza, "Twitter Bots Are Working to Suppress Voter Turnout to Stop Macedonia's NATO Membership: Report," *Newsweek*, 27 September 2018.

113 Metodieva, "How Disinformation Harmed the Referendum in Macedonia."

114 Karan, "#ElectionWatch: Boycott Campaign in Macedonia," DFRLab, Atlantic Council, 14 September 2018.

115 Sarah Bedenbaugh, Damon Wilson, and Graham Brookie, "Macedonia Vote Is Not the End of the Road," Atlantic Council, 3 October 2018.

116 Saska Cvetkovska, "Russian Businessman Behind Unrest in Macedonia," Organized Crime and Corruption Reporting Project (OCCRP), 16 July 2018.

117 Ibid.

118 "Macedonia Referendum: Name Change Vote Fails to Reach Threshold," *British Broadcasting Corporation*, 1 October 2018.

119 Elena Becatoros and Konstantin Testorides, "Macedonia: Referendum Approves Name Change, but Turnout Low," Associated Press, 30 September 2018.

120 Smith. "Macedonia Officially Changes Its Name to North Macedonia."

121 "About Truthmeter," Truthmeter.mk, *http://truthmeter.mk/about-truthmeter/*.

122 Karan, "#ElectionWatch: Boycott Campaign in Macedonia," Petreski, "#ElectionWatch: Fascist Falsification Ahead of Macedonian Referendum," Vladimir Radomirovic, "In the Balkans, Whistle-Blowing News Outlets Struggle to Survive," Nieman Reports, 17 March 2015; Sasha Cvetkovska, Aubrey Belford, Craig Silverman, J. Lester Feder, "The Secret Players Behind Macedonia's Fake News Sites," OCCRP, 18 July 2018.

123 Darko Brkan, telephone interview by Katie Joseff, 9 July 2019.

124 Erin Murrock, Joy Amulya, Mehri Druckman, and Tetiana Liubyva, *Winning the War on State-Sponsored Propaganda* (Washington, DC: International Research and Exchanges Board, 2018).

125 Mark Stencel, "Number of Fact-checking Outlets Surges to 188 in More Than 60 Countries," Poynter, 11 June 2019, *https://reporterslab.org/number-of-fact-checking-outlets-surges-to-188-in-more-than-60-countries/*.

126 Brendan Nyhan and Jason Reifler, "The Effect of Fact-checking on Elites: A Field Experiment on US State Legislators," *American Journal of Political Science* 59, no. 3 (2015): 628–640.

127 Jude Dineley, "Fact-checking—An Effective Weapon Against Misinformation?" Lindau Nobel Laureate Meetings, 5 April 2018, *www.lindau-nobel.org/blog-fact-checking-an-effective-weapon-against-misinformation/*.

128 Laura Hazard Owen, "Few People Are Actually Trapped in Filter Bubbles. Why Do They Like to Say That They Are?" Nieman Lab, 7 December 2018, *www.niemanlab.org/2018/12/few-people-are-actually-trapped-in-filter-bubbles-why-do-they-like-to-say-that-they-are/*.

129 Dineley, "Fact-checking—An Effective Weapon Against Misinformation?"

130 Thomas Wood and Ethan Porter, "The Elusive Backfire Effect: Mass Attitudes' Steadfast Factual Adherence," *Political Behavior* 41, no. 1 (2019): 135–163.

131 Dineley, "Fact-checking—An Effective Weapon Against Misinformation?"

132 Briony Swire, Adam J. Berinsky, Stephan Lewandowsky, and Ullrich K. H. Ecker, "Processing Political Misinformation: Comprehending the Trump Phenomenon," *Royal Society Open Science* 4, no. 3 (2017): 160802.

133 Adam J. Berinsky, ed., *New Directions in Public Opinion* (Routledge, 2015).

134 Nicholas A. Valentino, Vincent L. Hutchings, Antoine J. Banks, and Anne K. Davis, "Is a Worried Citizen a Good Citizen? Emotions, Political Information Seeking, and Learning Via The Internet," *Political Psychology* 29, no. 2 (2008): 247–273.

135 C. Shao, P. M. Hui, L. Wang, X. Jiang, A. Flammini, F. Menczer, and G. L. Ciampaglia, "Anatomy of an Online Misinformation Network," *PLOS ONE* 13, no. 4 (2018); J. Shin and K. Thorson, "Partisan Selective Sharing: The Biased Diffusion of Fact-checking Messages on Social Media," *Journal of Communication* 67, no. 2 (2017): 233–255; B. Nyhan, and J. Reifler, "Estimating Fact-checking Effects: Evidence from a Long-term Experiment During Campaign 2014," American Press Institute, 28 April 2015.

136 Renee Hobbs and Amy Jensen, "The Past, Present, and Future Of Media Literacy Education," *Journal of Media Literacy Education* 1, no. 1 (2009): 1.

137 Monica Bulger and Patrick Davison, *The Promises, Challenges, and Futures of Media Literacy* (New York: Data and Society Research Institute, 2018).

138 Ibid.

139 Ibid.

140 Se-Hoon Jeong, Hyunyi Cho, and Yoori Hwang, "Media Literacy Interventions: A Meta-analytic Review," *Journal of Communication* 62, no. 3 (2012): 454–472.

141 Bulger and Davison, *The Promises, Challenges, and Futures of Media Literacy*.

142 Sercan Ö. Arik, Mike Chrzanowski, Adam Coates, Gregory Diamos, Andrew Gibiansky, Yongguo Kang, Xian Li, and others, "Deep Voice: Real-time Neural Text-to-Speech," *ICML'17 Proceedings of the 34th International Conference on Machine Learning* 70 (2017): 196–204.

143 Derek Hawkins, "Reddit Bans 'Deepfakes,' Pornography Using the Faces of Celebrities such as Taylor Swift and Gal Gadot," *Washington Post*, 8 February 2018; Hans von der Burchard, "Belgian Socialist Party Circulates 'Deep Fake' Donald Trump Video," *Politico*, 21 May 2018.

144 "List of Deep Fake Tools," Vuild, 17 July 2019, *https://vuild.com/deep-fake-tools*.

145 Martijn Rasser, "Why Are Deepfakes So Effective?" *Scientific American*, 14 August 2019.

146 Timothy McLaughlin, "How WhatsApp Fuels Fake News and Violence in India," *Wired*, 12 December 2018.

147 Sam Gregory, "Deepfakes and Synthetic Media: What Should We Fear? What Can We Do?" Witness, 2018.

148 Julia Alexander, "YouTube Executives Reportedly Mulling Over Removing All Children's Content from Main Site," *Verge*, 19 June 2019.

149 Stephan Lewandowsky, Ulrich K. H. Ecker, Colleen M. Seifert, Norbert Schwarz, and John Cook, "Misinformation and Its Correction: Continued Influence and Successful Debiasing," *Psychological Science in the Public Interest* 13, no. 3 (2012): 106–131.

150 Drew Harwell, "Faked Pelosi Videos, Slowed to Make Her Appear Drunk, Spread Across Social Media," *Washington Post*, 24 May 2019.

151 Cuiha Shen, Mona Kasra, Wenjing Pan, Grace A. Bassett, Yining Malloch, and James F. O'Brien, "Fake Images: The Effects of Source, Intermediary, and Digital Media Literacy on Contextual Assessment of Image Credibility Online," *New Media & Society* 21, no. 2 (February 2019): 438–463.

152 "How Mass Surveillance Works in Xinjiang, China: Reverse Engineering Police App Reveals Profiling and Monitoring Strategies," New York: Human Rights Watch, 2 May 2019.

153 Joonas Rokka and Massimo Airoldi, "Cambridge Analytica's 'Secret' Psychographic Tool Is a Ghost from the Past," *The Conversation*, 2 April 2018.

154 Andrew McStay, "Now Advertising Billboards Can Read Your Emotions ... and That's Just the Start," *The Conversation*, 4 August 2015.

155 Dipayan Ghosh and Ben Scott, *#Digitaldeceit: The Technologies Behind Precision Propaganda on the Internet* (Washington, DC: New America, 2018).

156 Lisa-Maria Neudert, "Future Elections May Be Swayed by Intelligent, Weaponized Chatbots," *MIT Technology Review*, 22 August 2018.

157 Shawn Musgrave, "I Get Called a Russian Bot 50 Times a Day," *Politico*, 9 August 2017.

158 Adam D. I. Kramer, Jamie E. Guillory, and Jeffrey T. Hancock, "Experimental Evidence of Massive-Scale Emotional Contagion Through Social Networks," *Proceedings of the National Academy of Sciences* 111, no. 24 (2014): 8788–8790.

159 Aric Jenkins, "The Fall and Rise of VR: The Struggle to Make Virtual Reality Get Real," *Fortune*, 10 June 2019.

160 David Deal, "Virtual Reality Helps U.S. Athletes Train to Win Olympic Gold," *Medium*, 17 February 2018.

161 Luke Dormehl, "The U.S. Army Is Building a Giant VR Battlefield to Train Soldiers Virtually," Digital Trends, 20 March 2019.

162 Jenkins, "The Fall and Rise of VR: The Struggle to Make Virtual Reality Get Real."

163 Steve Dent, "Renault's Concept EV Drove Me at 80MPH While I Wore a VR Headset," Endgadget, 13 December 2017.

164 Jake Silverstein, "The Displaced: Introduction," *New York Times*, 5 November 2015.

165 Andrew Webster, "Pokémon Go's Wild First Year: A Timeline," *Verge*, 6 July 2017.

166 Anmar Frangoul, "Yankee Stadium soccer game gets the virtual reality treatment," CNBC, 25 July 2019.

167 Dominique Makowski, Marco Sperduti, Serge Nicolas, and Pascale Piolino, "'Being There' and Remembering It: Presence Improves Memory Encoding," *Consciousness and Cognition* 53 (2017): 194–202.

168 Kramer, Guillory, and Hancock, "Experimental Evidence of Massive-scale Emotional Contagion through Social Networks."

169 Giuseppe Riva, Fabrizia Mantovani, Claret Samantha Capideville, Alessandra Preziosa, Francesca Morganti, Daniela Villani, Andrea Gaggioli, Cristina Botella, and Mariano Alcañiz, "Affective Interactions Using Virtual Reality: The Link Between Presence and Emotions," *CyberPsychology & Behavior* 10, no. 1 (2007): 45–56.

170 Fiachra O'Brolcháin, Tim Jacquemard, David Monaghan, Noel O'Connor, Peter Novitzky, and Bert Gordijn, "The Convergence of Virtual Reality and Social Networks: Threats to Privacy and Autonomy," *Science and Engineering Ethics* 22, no. 1 (2016): 1–29.

171 Nexhmedin Morina, Hiske Ijntema, Katharina Meyerbröker, and Paul M. G. Emmelkamp, "Can Virtual Reality Exposure Therapy Gains Be Generalized to Real-Life? A Meta-analysis of Studies Applying Behavioral Assessments," *Behaviour Research and Therapy* 74 (2015): 18–24.

172 Andrew Guess, Jonathan Nagler, and Joshua Tucker, "Less Than You Think: Prevalence and Predictors of Fake News Dissemination on Facebook," *Science Advances* 5, no. 1 (2019).

PHOTO CREDITS

Cover photo – *iStock.com*/Eoneren; Ballot boxes – Octavio Hoyos/*Shutterstock.com*; Earthquake damage – *"Mexico City – Puebla 2017 Earthquake 3"* by *AntoFran* is licensed under *CC BY 4.0*; Map – Peter Hermes Furian/*Shutterstock.com*; VR headset – leungchopan/*Shutterstock.com*.

ABOUT THE AUTHORS

Samuel C. Woolley is a writer and researcher with a focus on emerging media technologies, propaganda and politics. He is also an assistant professor in the School of Journalism at the Moody College of Communication at the University of Texas-Austin and the program director of disinformation research at the University's Center for Media Engagement (CME).

His work looks at how automation, algorithms and AI are leveraged for both democracy and control. Prior to joining the faculty at UT, Woolley was the Director of Research of the ComProp Project at Oxford and the Director of the Digital Intelligence Lab at Institute for the Future (IFTF). He has served as a research fellow at Google Jigsaw, a resident fellow at the German Marshall Fund's Digital Innovation Democracy Initiative, a Belfer Fellow at the Center for Technology and Society at the Anti-Defamation League (ADL), a research fellow at the TechPolicy Lab at the University of Washington, and a pre-doctoral fellow at the Center for Media, Data and Society at Central European University.

His academic work has appeared in the *Journal of Information Technology and Politics,* the *International Journal of Communication,* the *Routledge Handbook of Media, Conflict and Security, A Networked Self: Platforms, Stories, Connections,* and *The Political Economy of Robots.* He has written for popular venues including *The Atlantic,* Wired, *The Guardian,* Motherboard-Vice, and Tech Crunch. His research has been featured in *The New York Times, The Washington Post,* and *The Wall Street Journal* and on Today, 60 Minutes, and Frontline. He holds a PhD from the University of Washington-Seattle.

Katie Joseff is the research manager of the Digital Intelligence Lab at Institute for the Future (IFTF), where she works to investigate computational propaganda—the use of automation, disinformation, and algorithms to manipulate public opinion online—and the ethical implications of emerging technologies.

At IFTF, she has conducted research on a wide variety of topics, including the "human consequences" of computational propaganda during the 2018 US midterm elections—a series of eight studies on impacted social groups and issue-publics (e.g. Jewish Americans, and immigration); the effects of false information on journalism and journalists; nuclear disinformation; psychological biases underlying propaganda; technology policy in the United States; and Ethical OS (a training series intended to illuminate future ethical risks related to technology).

She holds a BA and an MA from Stanford University, where she studied social neuroscience and international security as an undergraduate, and partisanship and disinformation as a master's student.

ABOUT THE FORUM

The International Forum for Democratic Studies at the National Endowment for Democracy (NED) is a leading center for analysis and discussion of the theory and practice of democracy around the world. The Forum complements NED's core mission—assisting civil society groups abroad in their efforts to foster and strengthen democracy—by linking the academic community with activists from across the globe. Through its multifaceted activities, the Forum responds to challenges facing countries around the world by analyzing opportunities for democratic transition, reform, and consolidation. The Forum pursues its goals through several interrelated initiatives: publishing the *Journal of Democracy,* the world's leading publication on the theory and practice of democracy; hosting fellowship programs for international democracy activists, journalists, and scholars; coordinating a global network of think tanks; and undertaking a diverse range of analytical initiatives to explore critical themes relating to democratic development.

ABOUT THE NATIONAL ENDOWMENT FOR DEMOCRACY

The National Endowment for Democracy (NED) is a private, nonprofit foundation dedicated to the growth and strengthening of democratic institutions around the world. Each year, NED makes more than 1,700 grants to support the projects of non-governmental groups abroad who are working for democratic goals in more than 90 countries. Since its founding in 1983, the Endowment has remained on the leading edge of democratic struggles everywhere, while evolving into a multifaceted institution that is a hub of activity, resources, and intellectual exchange for activists, practitioners, and scholars of democracy the world over.

ACKNOWLEDGEMENTS

The authors would like to thank Samantha Bradshaw, whose advice and feedback on an early draft of this paper led directly to improvements to the finished product's rigor and readability. Todd Helmus and Peter Kreko also provided valuable peer review, lending their professional expertise on specialized and technically complex subjects. The authors are grateful to Tyler Roylance for offering his outstanding editorial support.

Three staff members of the National Endowment for Democracy—Enrique Bravo-Escobar, Ivana Cvetković Bajrović, and Kaltrina Selmi—drew on a wealth of combined experience in Latin America and the Balkans to sharpen this paper's country examples. The authors also appreciate the contributions of the Forum's Shanthi Kalathil, Christopher Walker, Jessica Ludwig, Rachelle Faust, and Fabian Ringlund Hagemo. Particular acknowledgement goes to Dean Jackson, program officer for research and conferences with the International Forum, who served in a central role as the lead editor and coordinator of the overall production of the working paper.

National Endowment
for Democracy
Supporting freedom around the world

1025 F Street, NW, Suite 800 ■ Washington, DC 20004 ■ (202) 378-9700 ■ ned.org

ThinkDemocracy

@thinkdemocracy