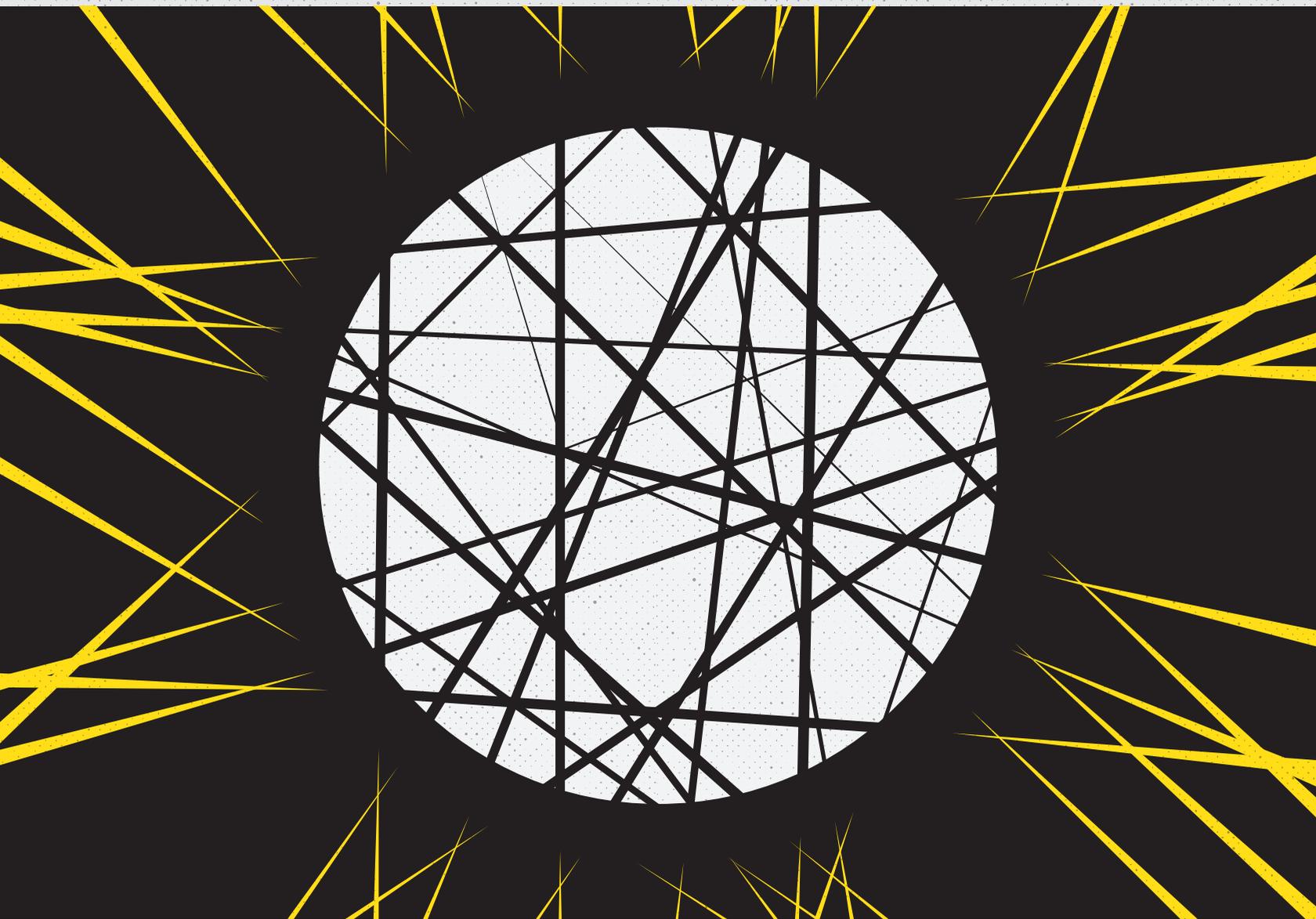


# UNDERSTANDING AND ADDRESSING THE DISINFORMATION ECOSYSTEM



**DECEMBER 15-16, 2017**

**ANNENBERG SCHOOL FOR COMMUNICATION**



**FIRST  
DRAFT**



THIS WORKSHOP BRINGS TOGETHER ACADEMICS, JOURNALISTS, FACT-CHECKERS, TECHNOLOGISTS, AND FUNDERS TO BETTER UNDERSTAND THE CHALLENGES PRODUCED BY THE CURRENT DISINFORMATION ECOSYSTEM. THE FACILITATED DISCUSSIONS WILL HIGHLIGHT RELEVANT RESEARCH, SHARE BEST-PRACTICES, IDENTIFY KEY QUESTIONS OF SCHOLARLY AND PRACTICAL CONCERN REGARDING THE NATURE AND IMPLICATIONS OF THE DISINFORMATION ECOSYSTEM, AND OUTLINE A POTENTIAL RESEARCH AGENDA DESIGNED TO ANSWER THESE QUESTIONS.

THIS CONFERENCE HAS BEEN MADE POSSIBLE THROUGH GENEROUS FUNDING FROM THE JOHN S. AND JAMES L. KNIGHT FOUNDATION.

# TABLE OF CONTENTS

## **Fake News in Context**

Michael Schudson (Columbia University)	
Barbie Zelizer (University of Pennsylvania)	1

## **Information Disorder: Definitions**

Hossein Derakhshan (MIT Media Lab and Harvard University Shorenstein Center)	
Claire Wardle (First Draft)	5

## **People Don't Trust News Media – And This Is Key to the Global Misinformation Debate**

Richard Fletcher (Reuters Institute for the Study of Journalism, University of Oxford)	
Rasmus Nielsen (Reuters Institute for the Study of Journalism, University of Oxford)	13

## **Taking the Red Pill: Ideological Motivations for Spreading Online Disinformation**

Rebecca Lewis (Data & Society Research Institute)	
Alice Marwick (University of North Carolina)	18

## **The Minority Report on the Fake News Crisis: (Spoiler Alert: It's the Real News)**

Duncan Watts (Microsoft Research)	
David Rothschild (Microsoft Research)	23

## **Personalized Information Environments and Their Potential Consequences for Disinformation**

Deen Freelon (University of North Carolina)	38
---	----

## **Making Sense of Information and Judging its Credibility**

Natalie Jomini Stroud (University of Texas)	
Emily Thorson (Syracuse University)	
Dannagal Young (University of Delaware)	45

## **Misinformation and Identity-Protective Cognition**

Dan Kahan (Yale University)	51
-----------------------------	----

## **Using Behavioral Theory to Curb Misinformation Sharing**

Brian Southwell (RTI International)	
Vanessa Boudeywns (RTI International)	58

## **Assessing Current Efforts by the Platforms and their Effectiveness**

Claire Wardle (FirstDraft)	64
----------------------------	----

## **Information Operations and Democracy: What Role for Global Civil Society and Independent Media?**

Sarah Oh (University of California, Berkeley)	69
---	----

## **Why We Feel Woefully Under-Informed About How We Can Do Our Jobs Better**

Amy Sippett (Full Fact)	75
-------------------------	----

## **Field Notes: Observation from a School Library**

Michael Barker (Phillips Academy)	80
-----------------------------------	----

## **How Academics Can Help Platforms Tackle Disinformation**

Nic Dias (FirstDraft)	83
-----------------------	----

## **Face2Face: Real-time Face Capture and Reenactment of RGB Videos**

Justis Theis (Technical University of Munich & University of Erlangen-Nuremberg)	
Michael Zollhöfer (Max-Planck Institute for Informatics)	
Marc Stamminger (University of Erlangen-Nuremberg)	
Christian Theobalt (Max-Planck Institute for Informatics)	
Matthias Nießner (Technical University of Munich)	87

## Fake News in Context

**Michael Schudson**, Columbia University  
**Barbie Zelizer**, University of Pennsylvania

“Fake news” landed in the US lexicon in 2016 as if it had arrived from Mars in a flying saucer, without precedent, without lineage. Fears about its influence were widespread, anxieties about its origins, both foreign and domestic, were large and alarm bells were clamorous about its threat to democracy. But now, with the initial panic at least partially in remission, we may be able to put it in better perspective.

In many ways, the noise created in response to fake news was predictable. From anti-Semitic blood libel stories in 15th century Europe to church-supported missives of divine retribution following the 1755 Lisbon Earthquake, fake news has been around a long while. Dating back hundreds of years to the invention of print and offering primarily inflammatory and sensational accounts of the events of the day, there has been a long history of journalistic hoaxes, scandals, lies, satire and exaggerations. In the US Thomas Jefferson complained in 1807 to a friend that “the man who never looks into a newspaper is better informed than he who reads them; inasmuch as he who knows nothing is nearer to truth than he whose mind is filled with falsehoods & errors” (Jefferson 1807). Fake news produced antebellum stories of African-Americans turning white or tales of alien populations on the moon, and it prevailed, according to Robert Darnton, in the *pasquinades* of 16<sup>th</sup> century Italy, the *canards* of 17<sup>th</sup> century Paris and the “paragraph men” and *nouvellistes* of London and Paris respectively who turned 18<sup>th</sup> century coffeehouse gossip into printed or handwritten papers left on public benches

for passersby, not unlike the rumor mongers of today’s digital environment (Darnton 2017).

Fake news lessened in centrality only in the late 1800s as printed news, particularly in Britain and the United States, came to center on what Jean Chalaby called “fact-centered discursive practices” and people realized that newspapers could compete with one another not simply on the basis of partisan affiliation or on the quality of philosophical and political essays but on the immediacy and accuracy of factual reports (Chalaby 1996). Until that point, few clear notions of journalistic ethics, balance or truth-telling were around to inspire trust in the reliability of news. In the 1890s early journalism textbooks could still recommend strict accuracy about the “essential” facts of a story while taking a tolerant attitude toward pure invention when it came to “non-essentials” (Shuman 1894) And fakery in photography and news photography early in the 20th century was widely accepted as just a part of the game (Tucher 2017).

So even then fake news persisted. It remained consonant with the sensationalism of the tabloids: The exaggerations and lies in the run-up to the Spanish-American War helped make yellow journalism profitable. But attempts to rein in fabrication appeared soon thereafter, notably through norms of media truthfulness and formal codes of ethics, beginning with the Kansas State Editorial Association in 1910. Reactions against Wilson administration propaganda to whip up support for World War I and

against the rapid rise of public relations after the war led more news organizations and associations of journalists to adopt ethical codes, taking shape nationally with the formation of the American Society of Newspaper Editors that in 1923 adopted a code of journalistic ethics (Samuel 2016). In the 1920s journalists adopted the ideal of “objectivity” in news reporting as a flag they would salute (Schudson 1978).

Fake news echoed around the world during World War I, where longstanding practices of propaganda and disinformation fostered the deliberate creation of false news by official or quasi-official agencies intent on destabilizing the public or shaping public opinion. Latent and manifest attempts at disinformation used both covert and conventional channels, making it difficult to know what was true or fake in an ecosystem that accommodated both: British disinformation efforts during World War II, capped by journalist Sefton Delmer’s so-called “Black Propaganda” campaign across Nazi territory, coexisted alongside Allied military efforts at publicizing true but hard-to-believe news of the concentration camps (Zelizer 1998). Fake news intensified during the Cold War, where the Russian practice of *dezinformatsiya* and similar US efforts vied with each to produce accounts damaging to the other side; portraying JFK’s assassination as a CIA plot was among them (Zelizer 1992). Today some argue that notions of propaganda aptly describe the contemporary information efforts of many online actors and activists (Jack 2017).

Fake news, then, has proliferated in various ways. No surprise that the fake news of one situation is not the fake news of another. Variations in detail, degree and intensity make the label a catch-all phrase, an imprecise moniker for the many kinds of

discredited accounts of purportedly reliable information.

Different forms of fake news easily spread across the media. Its adaptable shapes—where it can appear as honest error, misreporting, conspiracy theories, clickbait, unsubstantiated gossip, junk science and downright false news—underscore the hospitable nature of current circumstances, where its dissemination is readily facilitated by many features. They include algorithmic formulae that generate news feeds without regard for accuracy; a news environment driven not only by journalists; access to instantaneous and global dissemination; populist political currents that invoke the label of fake news as a weapon; low occupational esteem and failing business models across journalism; journalistic resistance to embracing perspectival forms of information favored by many beyond the news—blogs, podcasts, crowd-sourcing, listicles, gifs and memes; and populations who distrust the media while lacking the media literacy skills to discern or understand signs of journalistic bias, subtle agendas, or distinctions between news and advertisement. While today’s fake news builds on longstanding journalistic, technological, political, social, cultural, economic and legal contours that for the most part preceded the contemporary landscape, they nonetheless suit its current spread particularly well.

To act as if today’s fake news environment is fundamentally different from that of earlier times misreads how entrenched fake news and broader attitudes toward fakery have been. Today we readily accept fakery in many forms: reality television that is more staged than real, message manipulation and photoshopping in politics and entertainment, practices of appropriation that produce much-discussed art forgeries,

pirated computer software, musical remixes and designer fashion rip-offs. Fake news thus echoes an environment that is used to positioning falsehood as a seamless feature of public life.

Though the stubborn persistence of fake news across evolving and disparate circumstances is not new, three things strike us as different, at least in the U.S. context. First, there is great anxiety today about the border between professional journalists and others who through digital media have easy access to promoting their ideas, perspectives, factual reports, pranks, inanities, conspiracy theories, fakes and lies. Second, the President of the United States turns out to be either naïve about how to distinguish truth from fantasy or a faux naïf who effectively masquerades as a gullible believer in wild fantasies in order to draw attention away from or otherwise cast doubt on truthful reports that place him in a bad light. Finally, there is widespread recognition in the academy and beyond that getting at the truth is difficult, that the construction of what counts as truthful is indeed contestable and contested and that debunking falsehood requires critical skills and moral commitment to truth-seeking. The validity of this understanding that truths are socially constructed stops far short of saying “anything goes.”

What to do, then, about fake news? Currently there are admirable efforts to specify what fake news looks like, where and how it surfaces, what influence it has and how its distribution might be curtailed by watchdog journalists. If necessary, government regulation might curtail fake news, just as it does fraudulent advertising. But in the interim we need to keep clarifying how fake news is distinguishable from legitimate “fact-based discursive practices” and, equally, from openly partisan but

honest journalism. We also need to remember that fake news has a history. While that does not preclude recognizing its pernicious nature, at the least it should help us to keep it in perspective.

## References

- Chalaby, J. “Journalism as an Anglo-American Invention,” *European Journal of Communication* 11 (3), 1996, 303-326.
- Darnton, R. “The True History of Fake News,” *New York Review of Books*, February 13, 2017. Retrieved from <http://www.nybooks.com/daily/2017/02/13/the-true-history-of-fake-news/?printpage=true>
- Jack, C. “What’s Propaganda Got To Do With It,” *Data and Society: Points*. January 5, 2017. Retrieved from <https://points.datasociety.net/whats-propaganda-got-to-do-with-it-5b88d78c3282>
- Jefferson, T. Letter to John Norvell, June 11, 1807. In Adrienne Koch and William Peden, eds. *The Life and Selected Writings of Thomas Jefferson* (New York: Modern Library, 1944), 581.
- Samuel, A. “To Fix Fake News Look to Yellow Journalism,” *J-Stor Daily*, November 29, 2016. Retrieved from <https://daily.jstor.org/to-fix-fake-news-look-to-yellow-journalism/>
- Schudson, M. *Discovering the News: A Social History of American Newspapers* (New York: Basic Books, 1978).
- Shuman, E. L. *Steps Into Journalism* (Evanston, Ill: Correspondence School of Journalism, 1894), 66.

Tucher, A. "‘I Believe in Faking’: The Dilemma of Photographic Realism at the Dawn of Photojournalism," *Photography and Culture*, June 1 2017. Retrieved from <http://dx.doi.org/10.1080/17514517.2017.1322397>

Zelizer, B. *Covering the Body: The Kennedy Assassination, the Media and the Shaping of Collective Memory*. (Chicago: University of Chicago Press, 1992).

Zelizer, B. *Remembering to Forget: Holocaust Memory Through the Camera's Eye*. (Chicago: University of Chicago Press, 1998).

## Information Disorder: Definitions

**Hossein Derakhshan**, Massachusetts Institute of Technology Media Lab  
**Claire Wardle**, FirstDraft

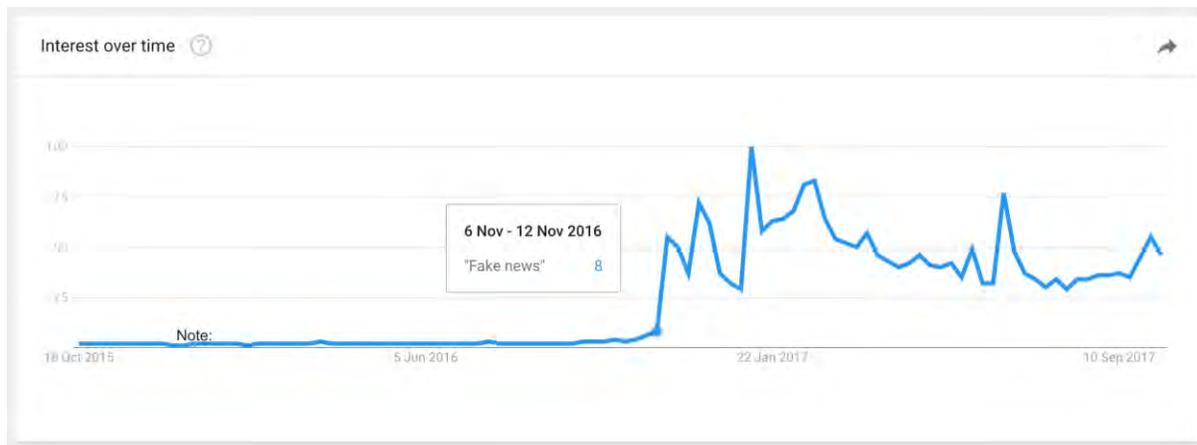
The use of deception and manipulation is as ancient as language itself, but according to Merriam Webster, the term ‘fake news’ only emerged around the end of the 19th Century. According to the dictionary website:

“One of the reasons that fake news is such a recent addition to our vocabulary is that the word fake is also fairly young. Fake was little used as an adjective prior to the late 18th century. [Before that point] the most common [description] was false news. We

can see this collocation in use as far back as the 16th century.”

The term appears new, because of its sudden appearance in our public discourse, caused mostly by its use by Donald Trump himself. This chart from Google Trends (a site that maps how frequently people search for a word or phrase on Google) shows that people suddenly started searching for that term between November 6 and 12, 2016. The US Presidential election was November 8.

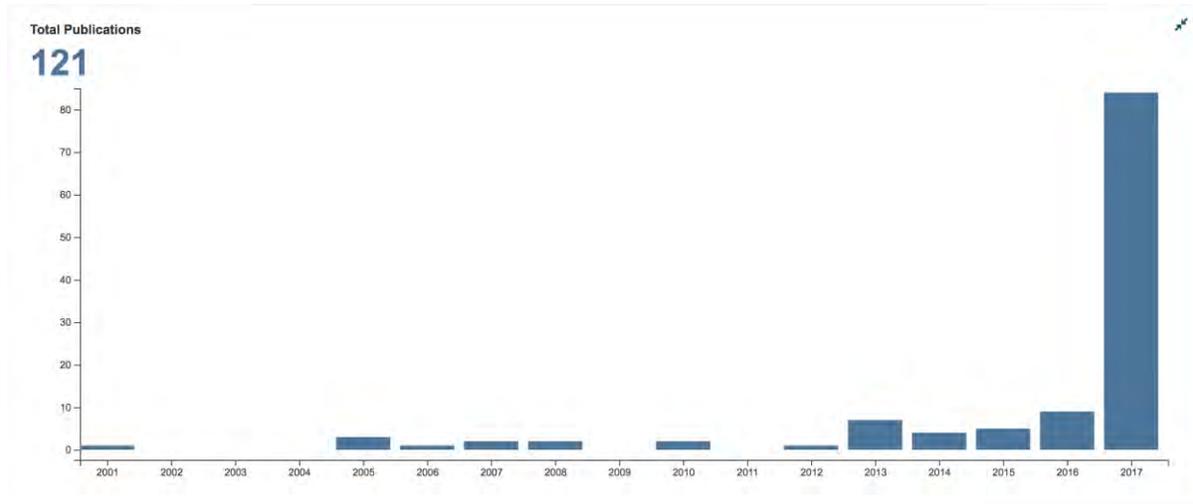
Figure 1: Google Trends data for the past 2 years, which shows the number of searches for the phrase ‘fake news.’



The academic community has followed this curve. This chart which visualizes peer-reviewed papers which included the terms

'fake news,' shows that this subject has become much more popular recently.

Figure 2: Number of peer-reviewed publications from the Web of Science that included ‘fake news’ in the title, abstract, or keywords.



In the months that have followed the US election, we’ve seen phrases such as alternative facts, post-truth, and post-fact enter public discourse, and most troubling, the term ‘fake news’ has been used as a weapon to discredit the media, undermining the concept of a free press. In this paper, we argue that while terminology and definitions always matter, on this topic, which is complex and sprawling, and involves information itself being used as a form of warfare, definitions are critical. Only by developing agreed upon definitions, and breaking down pieces of the ecosystem effectively, can we start to work collectively on solutions. In this paper, we introduce a new conceptual framework for understanding information disorder.

### The term ‘fake news’

One depressing aspect of the recent ‘fake news’ panic is that, while it has resulted in an astonishing number of reports, books, conferences, and events, it has produced little other than funding opportunities for research and the development of tools. One key reason for this stagnation, we argue, is

an absence of definitional rigor, which has resulted in a failure to recognize the diversity of mis- and dis-information, whether of form, motivation, or dissemination.

As researchers like Claire Wardle, Ethan Zuckerman, danah boyd, and Caroline Jack and journalists like the Washington Post’s Margaret Sullivan have argued, the term ‘fake news’ is woefully inadequate to describe the complex phenomena of mis- and dis-information. As Zuckerman states, “It’s a vague and ambiguous term that spans everything from false balance (actual news that doesn’t deserve our attention), propaganda (weaponized speech designed to support one party over another), and *disinformatzja* (information designed to sow doubt and increase mistrust in institutions).”

A study by Tandoc *et al.* published in August 2017 examined 34 academic articles that used the term “fake news” between 2003 and 2017. The authors noted that the term has been used to describe a number of different phenomena over the past 15 years: news satire, news parody, fabrication,

manipulation, advertising, and propaganda. Indeed, this term has a long history, long predating President Trump's recent obsession with the phrase.

The term "fake news" has also begun to be appropriated by politicians around the world to describe news organizations whose coverage they find disagreeable. In this way, it's becoming a mechanism by which the powerful can clamp down on, restrict, undermine, and circumvent the free press. It's also worth noting that the term and its visual derivatives (e.g., the red 'FAKE' stamp) have been even more widely appropriated by websites, organizations, and political figures identified as untrustworthy by fact-checkers to undermine opposing reporting and news organizations (Haigh 2017).

Many have offered new definitional frameworks in attempts to better reflect the complexities of mis- and dis-information.

Facebook defined a few helpful terms in their paper on information operations:

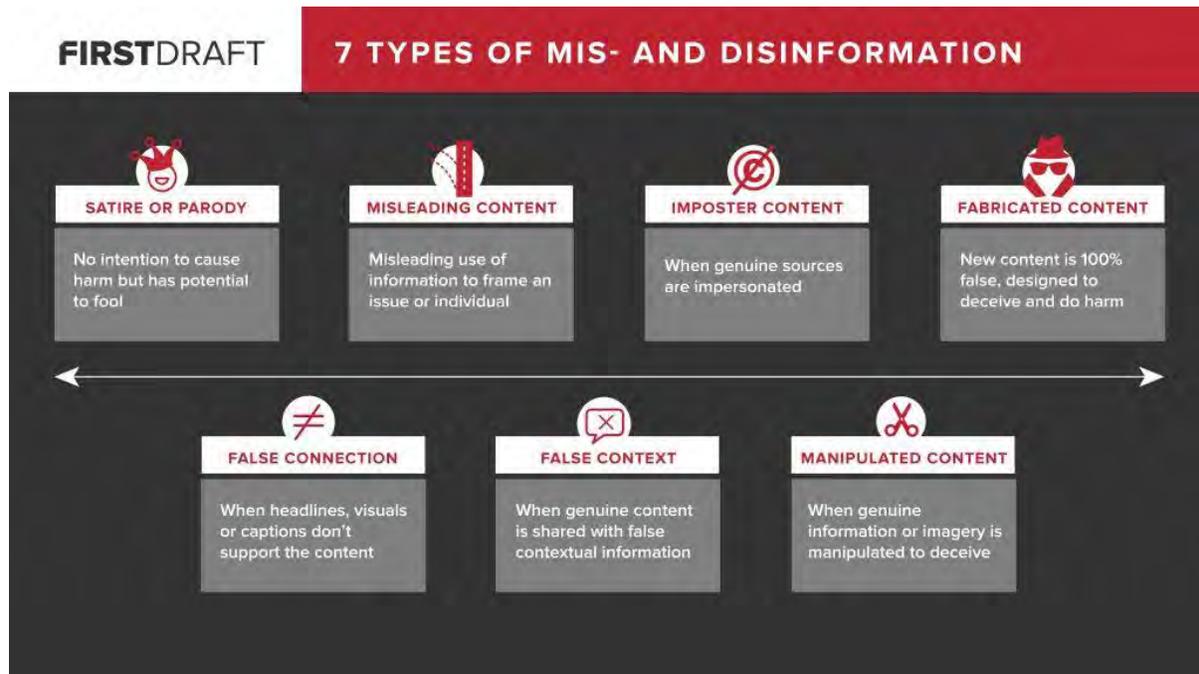
1. **Information (or Influence) Operations.** Actions taken by governments or organized non-state

actors to distort domestic or foreign political sentiment, most frequently to achieve a strategic and/or geopolitical outcome. These operations can use a combination of methods, such as false news, disinformation, or networks of fake accounts aimed at manipulating public opinion (false amplifiers).

2. **False News.** News articles that purport to be factual, but contain intentional misstatements of fact to arouse passions, attract viewership, or deceive.
3. **False Amplifiers.** Coordinated activity by inauthentic accounts that has the intent of manipulating political discussion (e.g., by discouraging specific parties from participating in discussion or amplifying sensationalistic voices over others).

In "Fake News. It's Complicated," Wardle outlines seven types of mis- and dis-information, revealing the wide spectrum of problematic content online, from satire and parody (which, while a form of art, can become misinformation when audiences misinterpret the message) to full-blown fabricated content.

Figure 3: 7 Categories of Information Disorder (Credit: Claire Wardle, First Draft)



While these seven classifications are helpful in encouraging people to see beyond the infamous ‘Pope endorses Trump’-type news sites that received so much attention after the US election, the phenomenon requires an even more nuanced conceptual framework—particularly one that highlights the impact of visuals in perpetuating disinformation. We have therefore created such a framework, and we will use it as the organizing structure for the report.

While we work through terms and descriptions, it’s important that we recognize the importance of shared definitions. As Caroline Jack argued in the introduction to her recent report, *Lexicon of Lies*, for Data & Society:

“Journalists, commentators, policymakers, and scholars have a variety of words at their disposal — propaganda, disinformation, misinformation, and so on — to describe the accuracy and relevance of media content.

These terms can carry a lot of baggage. They have each accrued different cultural associations and historical meanings, and they can take on different shades of meaning in different contexts. These differences may seem small, but they matter. The words we choose to describe media manipulation can lead to assumptions about how information spreads, who spreads it, and who receives it. These assumptions can shape what kinds of interventions or solutions seem desirable, appropriate, or even possible.”

Our conceptual framework has three components, each of which is also broken down into three parts:

1. The Three Types of Information Disorder: Dis-information, Mis-information, and Mal-information
2. The Three Phases of Information Disorder: Creation, Production, and Distribution

### 3. The Three Elements of Information Disorder: Agent, Message, and Interpreter

#### The Three Types of Information Disorder

Much of the discourse on ‘fake news’ conflates three notions: mis-information, dis-information, and mal-information. But it’s important to distinguish messages that are true from those that are false, and messages that are created, produced, or distributed by “agents” who intend to do harm from those that are not:

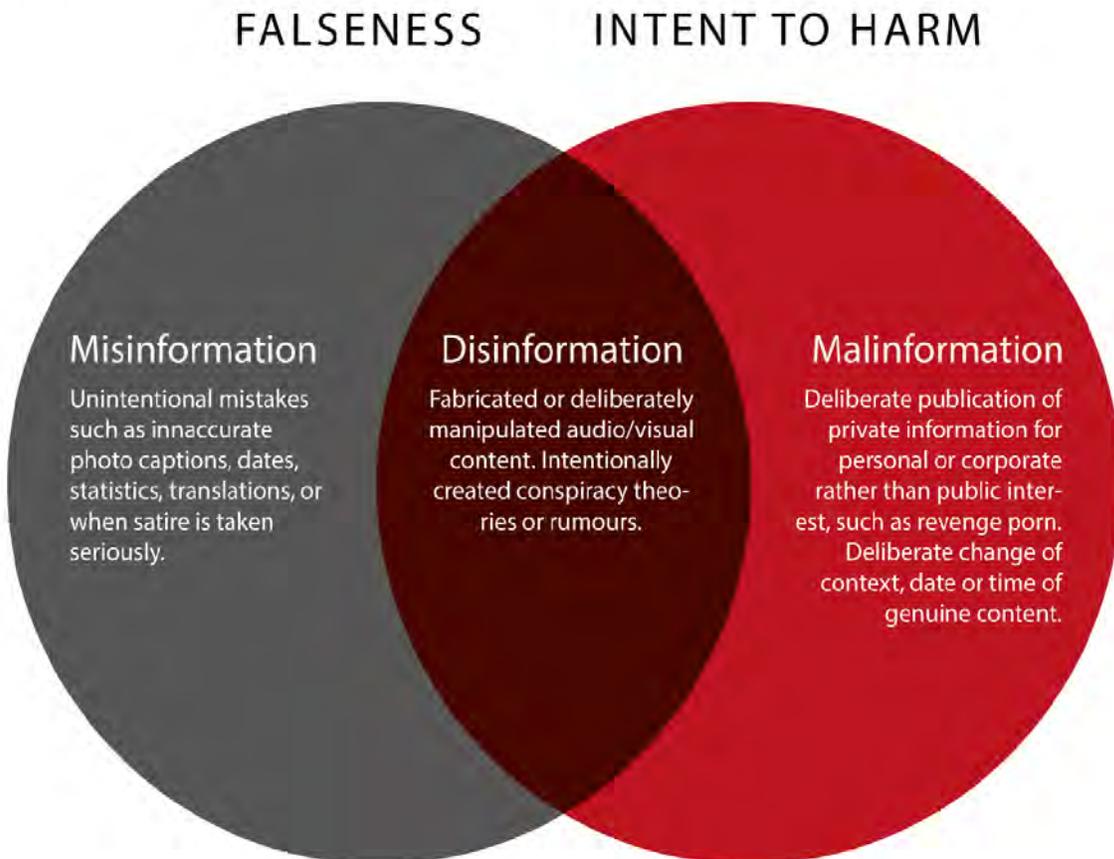
**Dis-information.** Information that is false and deliberately created to harm a person, social group, organization or country.

**Mis-information.** Information that is false, but not created with the intention of causing harm.

**Mal-information.** Information, that is based on reality, used to inflict harm on a person, organization, or country.

Figure 4: Examining how mis-, dis-, and mal-information intersect around the concepts of falseness and harm.

## TYPES OF INFORMATION DISORDER



**The Phases and Elements of Information Disorder**

In trying to understand any example of information disorder, it is useful to consider it in three elements:

1. **Agent.** Who were the ‘agents’ that created, produced, and distributed the example, and what was their motivation?

2. **Message.** What type of message was it? What format did it take? What were the characteristics?

3. **Interpreter.** When the message was received by someone, how did they interpret the message? What action, if any, did they take?

Figure 5: The Three Elements of Information Disorder



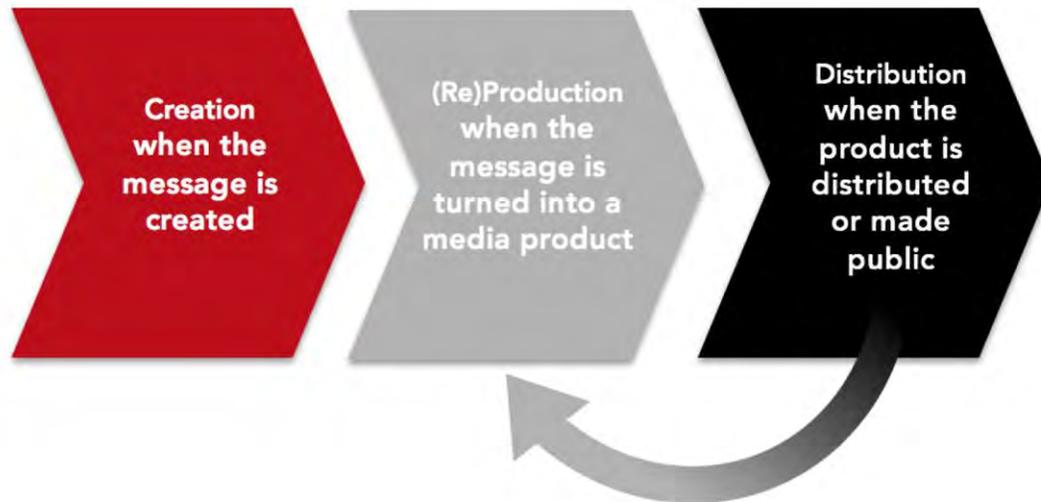
We argue that it is also productive to consider the life of an example of information disorder as having three phases:

1. **Creation.** The message is created.

2. **Production (and reproduction).** The message is turned into a media product.

3. **Distribution.** The message is distributed or made public.

Figure 6: The Three Phases of Information Disorder



In particular, it's important to consider the different phases of an instance of information disorder alongside its elements because the agent that creates the content is often fundamentally different from the agent who produces it. For example, the motivations of the mastermind who 'creates' a state-sponsored dis-information campaign are very different from those of the low-paid 'trolls' tasked with turning the campaign's themes into specific posts. And once a message has been distributed, it can be reproduced and redistributed endlessly, by many different agents, all with different motivations. For example, a social media post can be distributed by several communities, leading its message to be picked up and reproduced by the mainstream media and further distributed to still other communities.

In conclusion, only by dissecting information disorder in this manner can we begin to undertake the necessary research to fully understand the different facets of this phenomenon. And without methodologically rigorous, empirical research, we are certainly not going to understand the underlying symptoms that have caused the

current situation, or be in a position to work collaboratively on effective solutions.

#### Resources

boyd, d. (March 27, 2017) "Google and Facebook can't just make Fake News Disappear," *Wired*, <https://www.wired.com/2017/03/google-and-facebook-cant-just-make-fake-news-disappear/>

Haigh et al, (2017) "Stopping Fake News: The work practices of peer-to-peer counter propaganda." *Journalism Studies*, 1-26.

Jack, C. (2017) "Lexicon of Lies," *Data & Society*, [https://datasociety.net/pubs/oh/DataAndSociety\\_LexiconofLies.pdf](https://datasociety.net/pubs/oh/DataAndSociety_LexiconofLies.pdf)

Merriam Webster. "The Real Story of Fake News," <https://www.merriam-webster.com/words-at-play/the-real-story-of-fake-news>

Sullivan, M (Jan 6, 2017) "It's Time To Retire the Tainted Term Fake News," *Washington Post*,

[https://www.washingtonpost.com/lifestyle/style/its-time-to-retire-the-tainted-term-fake-news/2017/01/06/a5a7516c-d375-11e6-945a-76f69a399dd5\\_story.html](https://www.washingtonpost.com/lifestyle/style/its-time-to-retire-the-tainted-term-fake-news/2017/01/06/a5a7516c-d375-11e6-945a-76f69a399dd5_story.html)

Tandoc, Jr., E. C, Lim, Z. W., and Ling, R. (Aug. 2017) “Defining ‘Fake News’: A Typology of Scholarly Definitions,” *Digital Journalism*, 5 (7): 1-17

Wardle, C. (Feb 16, 2017) “Fake News. It’s Complicated,” *First Draft*, <https://firstdraftnews.com/fake-news-complicated/>

Zuckerman, E. (Jan 30. 2017) “Stop Saying Fake News, It’s not Helping,” *My Heart is in Accra*, <http://www.ethanzuckerman.com/blog/2017/01/30/stop-saying-fake-news-its-not-helping/>

*\*Some elements of this paper are taken from a report commissioned by the Council of Europe entitled “Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking.”*

## People Don't Trust News Media – and This is Key to the Global Misinformation Debate

**Richard Fletcher**, Reuters Institute for the Study of Journalism, University of Oxford  
**Rasmus Kleis Nielsen**, Reuters Institute for the Study of Journalism, University of Oxford

If we want to understand the global debate around misinformation, we have to accept that it is playing out in the context of low trust in news media, general scepticism of publishers and platforms alike, and in an environment where people do not make a clear-cut distinction between “real news,” fake news, and other forms of misinformation, but rather see the difference as one of degree. These are three key takeaways from our recent research into how people see news, media, and journalism across the world. There are significant differences from country to country, but also clear common patterns. Crudely put, people often have little trust in the news media, they are skeptical of the information they come across on platforms, and they see poor journalism, political propaganda, and misleading forms of advertising and click-bait as contributing to the problem.

### Data

To illustrate these three points, we turn to data from the 2017 Reuters Institute Digital News Report (Newman et al. 2017). The findings in the report are based on an online survey of around 70,000 respondents across 36 markets, as well as an additional eight focus group sessions in four of those (Finland, Spain, the UK, and the US). Polling was conducted by YouGov, and the focus groups were moderated by Kantar Media, but both the questionnaire and focus group discussion guides were designed by researchers at the Reuters Institute (including the authors). The survey used nationally-representative samples of around

2,000 in each country, with respondents selected based on interlocking quotas for age, gender, and location. The focus groups are not representative in a statistical sense, but were organized by age and were varied in terms of other demographic and news use characteristics (see Vir and Hall 2017 for more details). In an attempt to avoid generalizations based on extrapolating from single-country findings, we will focus here on 10 countries that together represent a wide range of different high-income democracies: Germany, France, Spain, Ireland, Denmark, Greece, Australia, Canada, the UK, and the US.

### First, trust in the news media is low

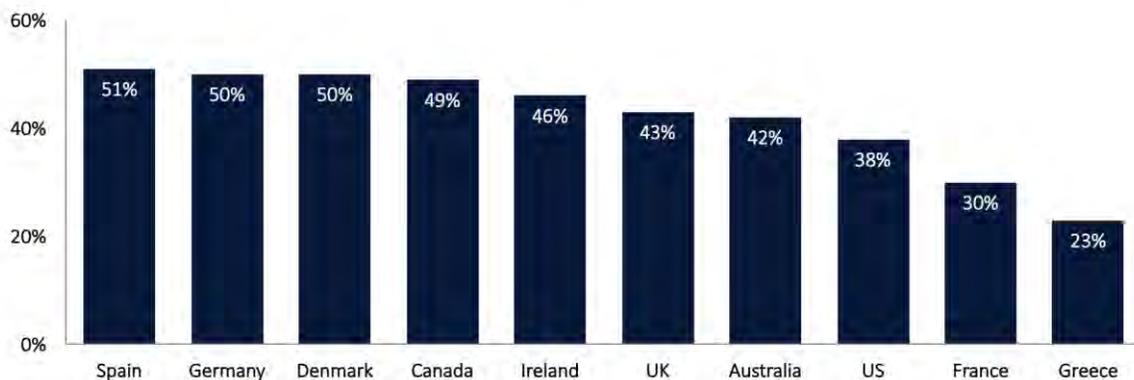
The first of our three points is that the debate about misinformation is often taking place against the backdrop of low levels of trust in the news media. As is clear from Figure 1, less than half in many countries agree that they can “trust most news most of the time.” Trust is particularly low in Greece (23%) and France (30%), but also in the US (38%). At most, this is only partly due to the current debates about misinformation, as we know from a handful of more detailed studies that in many countries—particularly the US (Ladd 2012)—trust in news media has been declining for years, as part of a broader crisis of confidence in many public institutions (Norris 2011). But it is also clear that this same fall in trust is not happening everywhere (Hanitzsch, van Dalen, and Steindl 2017). We can see from our own survey data that trust in the news is more widespread in Spain (51%), Germany

(50%), and Denmark (50%), with the highest figure of all in Finland (62%, but not considered here).

Trust is a complicated issue, and these figures can be interpreted in a number of ways. However, they should not be treated as a proxy measure for how trustworthy the news media actually are in certain countries. People might trust things that appear deeply untrustworthy from an outside or neutral perspective, particularly if they are unaware of potential alternatives, or if they happen to support the views and actions of the untrustworthy actors. Nor should the figures be used to straightforwardly understand news consumption, as we have known for

some time that some people regularly consume news they say they do not trust (Tsfati and Cappella 2003). However, this measure does give us an indication of people's attitudes towards the news media, and gives us some basis for thinking about how initiatives aiming to combat misinformation—particularly those that hinge on trust—might play out in particular contexts. Tagging social media posts as 'verified' by established news organizations might work well in Germany, where many trust the media, but perhaps less so in Greece, where most people—perhaps rightly—do not.

Figure 1. Proportion that 'strongly agree' or 'tend to agree' they can trust most news most of the time



## Second, confidence in social media is even lower

We mention social media because many see it as absolutely central to the debates about misinformation. We therefore also asked respondents a follow-up question about their confidence in the ability of both news media and social media to separate fact from fiction. We use it to make our second point, namely that although confidence in the news media is generally low, confidence in social media is even lower (see Figure 2). This is true of every country apart from Greece, which as we have already seen, has particularly low trust in the news media, but

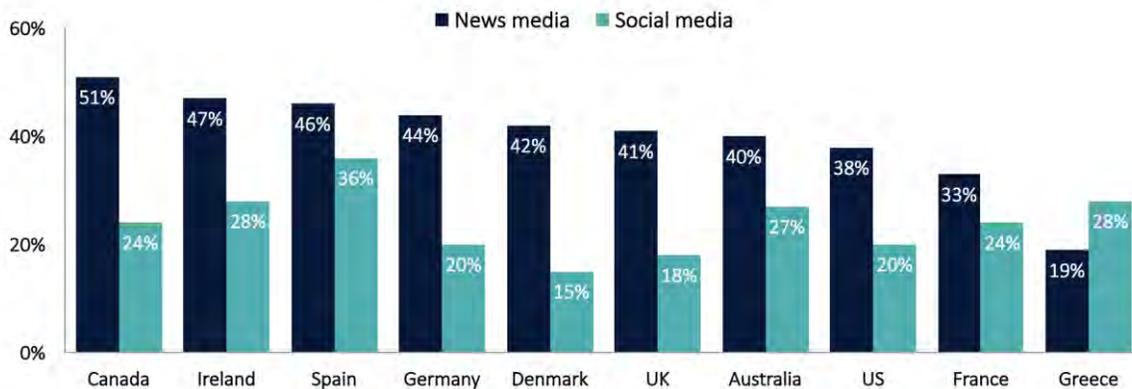
broadly comparable levels of confidence in social media. In the UK and the US, no more than one in five agree that social media does a good job in helping them distinguish fact from fiction. Of course, unlike the news media, few social networks claim that this is what they set out to do, but it is nonetheless striking that most people do not feel that social media help them much in this regard.

As people increasingly rely on social media to learn about what's happening in the world, we could interpret this state of affairs in quite a pessimistic fashion—as a catalyst for already-declining trust. But a more

optimistic reading reminds us that we should not conflate exposure to misinformation with uncritical acceptance of misinformation. Elsewhere, we have called the combination of low trust in news media and even lower confidence in social media

“generalized scepticism,” as people’s views of one are strongly correlated with their views of the other (Fletcher and Nielsen 2017).

Figure 2. Proportion that ‘strongly agree’ or ‘tend to agree’ that the news media/social media does a good job in helping them distinguish fact from fiction



### Third, many see fake news as different from poor journalism primarily by degree

In this environment of low trust in news media and even lower confidence in social media and much of the information accessed there, how do people think about the phenomenon of fake news? Top-down definitions abound, including useful typologies of misinformation (e.g. Wardle 2017) and calls for the term “fake news” to be abandoned altogether in favour of more precise vocabulary (Sullivan 2017). But our qualitative research demonstrates that the term is already part of the vernacular because it helps people express their frustration with the media environment, because it resonates with their experiences of coming across many different kinds of misinformation, especially online, and because it is actively used by critics of both news media and platform companies.

We have gathered qualitative evidence to provide an audience-centric, bottom-up set

of definitions of fake news from a subset of four countries, specifically the US and the UK (countries where less than half the population trust most news most of the time) and Finland and Spain (where just over half the population say they do so). The central point that came out of our focus groups was that most people do not operate with a categorical distinction between “fake news” and “real news,” but see the difference as one of degree, as this exchange from the UK shows:

Moderator: What does “fake news” mean to you?

M: Made up stories.

F: Do you believe everything that you hear and see and read? I’m sure some of it is made up.

M: It’s a spectrum isn’t it?

M: There’s been fake news for years hasn’t there?

We cannot present the focus group data in detail here, or unfold a full analysis (see

Nielsen and Graves 2017 for this). But Figure 3 provides a visual summary of the main types of content that people identified as fake news during the discussions, including poor journalism, political

propaganda, and misleading forms of advertising and clickbait. Satire and outright false news were also mentioned, but often explicitly discounted as “not news” at all.

Figure 3. Types of misinformation identified by our focus group participants and their overlap with news



This suggests how the global misinformation debate is playing out—not only against the current backdrop of low trust in news and generalized skepticism towards both publishers and platforms, but also against what is often years of frustrating experiences with many kinds of news.

**Conclusion**

In this essay, we have argued that the global debate around misinformation plays out in a context where people do not trust news media, they are skeptical of the information they come across on platforms, and they see poor journalism, political propaganda, and misleading forms of advertising and clickbait as contributing to the problem. There are very clear and important country-to-country differences in how pronounced these trends are. Trust in news media is significantly higher in Germany than in the

US, people have more confidence in social media in Spain than in the UK, and the fake news discussion is seen as much more relevant by focus group participants in English-speaking countries than elsewhere. But there are also clear commonalities. It is striking that even in relatively high-trust contexts, about half the population do not think they can trust most news most of the time. Whether this is indicative of healthy skepticism or a paralyzing form of hardened cynicism is a matter of judgement. Broadly speaking, societies benefit from people trusting trustworthy institutions and distrusting untrustworthy ones (O’Neill 2002). What we can say here is that, empirically, many people do not see publishers and platforms (or politicians for that matter) as particularly trustworthy when it comes to news and factual information.

This is key to the global misinformation debate.

## References

- Fletcher, Richard, and Rasmus Kleis Nielsen. 2017. "Navigating News on Social Media: A Four-Country Mixed-Methods Analysis." Paper presented at APSA Annual Meeting.
- Hanitzsch, Thomas, Arjen van Dalen, and Nina Steindel. 2017. "Caught in the Nexus: A Comparative and Longitudinal Analysis of Public Trust in the Press." *International Journal of Press/Politics* 23 (1):3–23.
- Ladd, Jonathan M. 2012. *Why Americans Hate the Media and How It Matters*. Princeton: Princeton University Press.
- Newman, Nic, Richard Fletcher, Antonis Kalogeropoulos, David A. L. Levy, and Rasmus Kleis Nielsen. 2017. "Reuters Institute Digital News Report 2017." Oxford: Reuters Institute for the Study of Journalism.
- Nielsen, Rasmus Kleis and Lucas Graves. 2017. "News You Don't Believe": Audience Perspectives on Fake News. Oxford: Reuters Institute for the Study of Journalism.
- Norris, Pippa. 2011. *Democratic Deficit: Critical Citizens Revisited*. Cambridge: Cambridge University Press.
- O'Neill, Onora. 2002. *A Question of Trust*. Cambridge: Cambridge University Press.
- Sullivan, Margaret. 2017. "It's Time to Retire the Tainted Term 'Fake News.'" *Washington Post*. [https://www.washingtonpost.com/lifestyle/style/its-time-to-retire-the-tainted-term-fake-news/2017/01/06/a5a7516c-d375-11e6-945a-76f69a399dd5\\_story.html](https://www.washingtonpost.com/lifestyle/style/its-time-to-retire-the-tainted-term-fake-news/2017/01/06/a5a7516c-d375-11e6-945a-76f69a399dd5_story.html).
- Tsfati, Yariv, and Joseph N. Cappella. 2003. "Do People Watch What They Do Not Trust? Exploring the Association Between News Media Skepticism and Exposure." *Communication Research* 30 (5):504–29.
- Vir, Jason, and Kathryn Hall. 2017. "Attitudes to Paying for Online News." Oxford: Reuters Institute for the Study of Journalism.
- Wardle, Claire. 2017. "Fake News. It's Complicated." *First Draft News*. <https://firstdraftnews.com:443/fake-news-complicated/>.

## Taking the Red Pill: Ideological Motivations for Spreading Online Disinformation

Rebecca Lewis, Data & Society Research Institute  
Alice Marwick, University of North Carolina at Chapel Hill

**Abstract:** This paper addresses the ideological motivations of members of far-right internet subcultures who spread misinformation and disinformation online. These groups recruit like-minded individuals who believe that white, male identities and conservative belief systems are under attack from a range of outside forces. They often greatly distrust the mainstream media, which they believe to be dominated by liberal ideology and “politically correct” culture. As group members are radicalized – a process they refer to as “redpilling” – their ideologies and distrust of the media feed on each other and ultimately inform a broader shift in their understanding of reality and veracity. As a result, they may view highly ideological and factually incorrect information as truthful, thus complicating understandings of disinformation.

The role of “fake news,” misinformation, and propaganda in spreading inaccurate information online came to prominence during the run-up to the 2016 election and remains an area of wide concern. Previous research has identified a variety of motivations behind such content, including financial interests (for instance, hyper-partisan clickbait content that promotes audience engagement), the political interests of state actors (such as computational propaganda spread by Russian actors on platforms like Twitter), trolling and disruption (for example, imageboard denizens who spread wild rumors for entertainment), and the desire for fame (as influencers like Milo Yiannopolous and Mike Cernovich came to prominence during GamerGate and have increased their online footprints by adopting alt-right beliefs) (Caplan & boyd, in press; Woolley & Howard, 2016). For the last year and a half, our team at Data & Society has conducted research into the manipulation of the mainstream media by far-right internet subcultures, including white supremacists, men’s rights activists, 4chan and 8chan

users, trolls, and conspiracy theorists (Marwick & Lewis, 2017).

For those on the far-right, recruiting like-minded individuals and converting them to their way of thinking is a primary driver of online participation. While these groups have robust online communities, blogs, podcasts and the like, the mainstream media offers a significant opportunity to set agendas, frame current issues, and spread their messaging to a larger and more diverse audience. Since these groups often adhere to conspiratorial thinking which maintains that white, male identities and conservative belief systems are under attack from a hostile Other, whether that be people of color, Muslim immigrants, Jews, global elites, feminists, or the Illuminati, they share a sense of urgency and significance in growing their movement. At the same time, their ideological presuppositions justify what is often aggressive online and offline harassment of feminists, people of color, Jews, journalists, and so forth by presenting an alternate reality in which white men are the victims of an aggressively politically-

correct liberal culture. While individuals may have a variety of reasons to participate in these subcultures, this essay discusses the ideological motivations behind these groups' spread of misinformation and disinformation online—that is, information that is incorrect, and information that is deliberately misleading (Jack, 2017).

### **Distrust of the Media**

Generally, far-right groups spreading mis- and disinformation have very low levels of trust in the mainstream media. To a certain extent, this is reflective of a larger trend; distrust in the media is at an all-time low nationwide, particularly among Republicans, where trust has dipped to 14% (Swift, 2016). Many right-leaning media consumers and young men more generally express frustration with the media's "PC culture," which they interpret as liberal bias, pointing to media ownership by liberal philanthropists like Jeff Bezos and Ted Turner as proof.

Far-right critiques of mainstream media go beyond perceived liberal bias. For years, self-described "trolls" have criticized the mainstream media for its overly sensational and emotionally manipulative tactics surrounding issues like pedophilia and the murders of young white women (Phillips, 2013). Separately, media consolidation at the corporate level has left "news deserts" in rural areas, meaning that rural news consumers may feel that local media fails to represent their interests. At a time when social media has removed gatekeepers and allowed anyone with internet access to express their opinions, news outlets represent a more traditional, closed system of information distribution, which many members of far-right movements characterize as a distant elite. The extreme version of this belief system maintains that the media deliberately pulls the wool over

the eyes of ordinary people in order to justify elite dominance, and often veers into familiar conspiracy theories about Jewish control. Similarly, the liberal beliefs of Silicon Valley media titans have increased distrust in social media platforms, especially when far-right participants believe that content moderation and trending topics favor left-wing viewpoints.

### **Redpilling and the Shifting Nature of Truth**

The process of "redpilling"—a shorthand for far-right radicalization—illustrates how far-right ideology and media distrust feed on each other and ultimately represent a broader shift in people's perceptions of truth, disinformation, and propaganda. The term *redpilling* speaks to a kind of total revelation, a "waking up" to the realities of the world. The "red pill" refers to the scene in *The Matrix* in which the character Morpheus offers the protagonist Neo the option of taking a red pill or a blue pill. By taking the red pill, Neo sees "the truth" about the Matrix and his world is upended; had he taken the blue pill, he would have returned to his everyday life, none the wiser. In far-right circles, one is redpilled when they begin believing in a truth that is counterfactual to mainstream belief, which may include white supremacy, Holocaust denial, the danger that immigration posits for white Americans, the oppression of men by feminists, and so forth.

People often first become redpilled in one ideological area, which then serves as an entry point to redpilling in other areas. In other words, our research indicates that someone who has been redpilled on one issue is more likely to be redpilled on others. In this way, susceptibility to redpilling resembles susceptibility to conspiracy theories, as research has shown that one of the greatest predictors of belief in a

conspiracy theory is belief in others (Douglas & Sutton, 2008). Most frequently, the world of Men's Rights activism—fueled by a backlash to popular feminism, or a lack of romantic or sexual success—serves as an entry point for disaffected young men into white nationalism. Significantly, far-right groups attempting to radicalize people may encourage this piecemeal approach. On the white supremacist blog *Fash the Nation*, for instance, a series called “Red Pill 101” includes articles on “equality,” “race,” and “African Americans,” but also on “pornography,” which they believe Jews control and use as a strategy to make white men impotent.

The example above shows not only how redpilling leads to a broader radicalization, but also how quickly ideology and distrust of the media feed on each other throughout the process. This also explains the far-right's proclivity for conspiracy theories. If someone begins to believe that white people are actually the oppressed race, it also means that everything that mainstream media is telling them is false. In white supremacist circles, it is widely believed that Jewish elites control the culture (specifically, the media) and are accelerating the destruction of the white race. So, as one becomes redpilled on white supremacy, he is also necessarily redpilled on the mainstream media. The more someone is radicalized against the mainstream media, the more the explanations of ideologues make sense to explain its shortcomings. Thus, redpilled individuals by their very nature distrust any “official story” given to them by mainstream media.

As a result, those who are redpilled fully reject and antagonize the mainstream media and its narratives, as well as traditional institutions of government, business, and education. Not only do the ideologies of the

newly-radicalized shift, but their very definitions of misinformation, disinformation, and propaganda, as well as concepts of oppression and inclusion, do as well. Redpilling thus begins to inform an entirely different view of reality, epistemology, and veracity. Redpilled individuals may not view the content that they spread as mis or disinformation (although in many cases they strategically re-frame events to encourage adoption of their own belief systems), but as necessary in order to protect their own interests and fight for the survival of their very identities.

### **Misinformation vs. Disinformation**

In January 2016, four Chicago teens used Facebook Live to film themselves torturing a mentally disabled man, which received widespread media attention. Alt-right media personality, author, and blogger Mike Cernovich used Twitter's streaming app Periscope to talk to his followers, brainstorming a way to connect the event to larger ideological goals. Since the teenagers were African-American, Cernovich and his followers decided to link the kidnapers to Black Lives Matter, a movement for racial justice which the far-right often describes as a terrorist organization. (One goal of the far-right is getting the federal government to officially label BLM and movements like Antifa terrorist organizations, which would curtail their activities and justify federal action against members.) Using bots and fake accounts, Cernovich and his followers tweeted the hashtag #BLMKidnapping 480,000 times in 24 hours, causing it to trend on Twitter. While both Chicago police and BLM activists debunked the rumor, it spread widely on social media and was picked up by hyper-partisan news sites like Breitbart. The idea that the kidnapers were BLM activists became so pervasive that it was mentioned in most media stories about the kidnapping, even if just to discredit it.

Cernovich and his audience forcibly changed the way the story was framed by pushing out the hashtag. We consider this an example of *disinformation*, since Cernovich and his audience knew that BLM was not actually linked to the Chicago kidnapping. Instead, they propagated an alternate frame in order to further their larger ideological goals.

By contrast, consider how the far-right has propagated information about Antifa. Antifa is a term for antifascist groups who use direct action techniques to protest fascist and neo-Nazi thought, often through physical force. In the US, this has been a very small group until quite recently, made up primarily of college-age young people, often linked to anarchist and punk rock communities. In the current political climate, where far-right groups are more visible than in the last two decades, Antifa activism is growing in numbers.

The YouTube documentary “America Under Siege: Antifa” frames Antifa as a radical movement opposed not just to fascism, but to all mainstream conservatism. It presents both accurate historical information about Antifa’s European roots and complete inaccuracies, such as the idea that Antifa controls entire towns in East Germany and prevents any right-wing beliefs from being publicly discussed, or that Antifa is funded by George Soros. Some of the documentary’s claims are ideologically-slanted, but not actually incorrect; for instance, the claim that Antifa members dislike anyone who espouses “Western free enterprise system capitalism,” can be read as accurate in one sense, given that Antifa activists are primarily anarchists who by definition are opposed to capitalism. In another sense it is entirely *inaccurate* – Antifa activists do not, generally, attack “anyone” who supports the status quo. This

murkiness, and the lack of understanding around intentionality, makes it difficult to determine whether this documentary is strategic disinformation. Instead, it might be considered *misinformation*, inaccurate information spread unintentionally. If its creators and distributors are redpilled, they may be entirely convinced that their interpretation of Antifa is correct.

## References

- Caplan, R., & boyd, danah. (in press). Who’s Playing Who? Media Manipulation in an Era of Trump. In Z. Papacharissi & P. J. Boczkowski (Eds.), *Trump and the Media*. Cambridge, MA: MIT Press.
- Douglas, K. M., & Sutton, R. M. (2008). The hidden impact of conspiracy theories: Perceived and actual influence of theories surrounding the death of Princess Diana. *The Journal of Social Psychology, 148*(2), 210–222.
- Jack, C. (2017). *Lexicon of Lies: Terms for Problematic Information*. New York, N.Y: Data & Society Research Institute. Retrieved from <https://datasociety.net/output/lexicon-of-lies/>
- Marwick, A., & Lewis, R. (2017). *Media Manipulation and Disinformation Online* (p. 107). New York: Data & Society Research Institute. Retrieved from <https://datasociety.net/output/media-manipulation-and-disinfo-online/>
- Phillips, W. (2013). The house that fox built: Anonymous, spectacle, and cycles of amplification. *Television & New Media, 14*(6), 494–509.

Swift, A. (2016, September 14). Americans' Trust in Mass Media Sinks to New Low. Retrieved from <http://www.gallup.com/poll/195542/americans-trust-mass-media-sinks-new-low.aspx>

Woolley, S. C., & Howard, P. N. (2016). Automation, Algorithms, and Politics| Political Communication, Computational Propaganda, and Autonomous Agents — Introduction. *International Journal of Communication*, 10(0), 9.

## **The Minority Report on the Fake News Crisis: (Spoiler Alert: It's the Real News)**

**Duncan J. Watts**, Microsoft Research  
**David M. Rothschild**, Microsoft Research

*"To ensure that the concept of self-government outlined by the U.S. Constitution remains a reality into future centuries, the American people must be well informed in order to make decisions regarding their lives, and their local and national communities. It is the role of journalists to provide this information in an accurate, comprehensive, timely and understandable manner."* –Society for Professional Journalists

### **Introduction**

Since the 2016 Presidential election, an increasingly familiar narrative has emerged concerning the unexpected victory of Donald Trump. Fake news, much of it produced by Russian sources, was amplified on social networks such as Facebook and Twitter, generating millions of views among a segment of the electorate eager to hear stories about Hilary Clinton's malfeasance and Obama's illegitimacy (Madrigal 2017; Palma and Green 2017; Roose 2017; Timberg 2017). "Alt-right" news sites like Breitbart and Daily Caller amplified and supplemented the outright fake information with highly slanted and misleading coverage of their own (Benkler et al. 2017a; Rutenberg 2017). In parallel, the continuing fragmentation of the media and the increasing ability of Americans to self-select into like-minded "filter bubbles" exacerbated both phenomena, generating a toxic brew of political polarization and skepticism toward traditional sources of authority (Delaney 2017; El-Bermaway 2016). Finally, powerful machine learning algorithms, coupled with reams of data from social networks, allowed groups allied with the Trump campaign to micro-target susceptible voters in swing states with

messages specifically tailored to shift their votes (Albright 2017).

Alarmed by these events, the mainstream media has vigorously shouldered the mantle of truth-speakers. The Washington Post changed its motto to "Democracy dies in Darkness" on February 22, 2017 (Fahri 2017). The New York Times launched a major ad campaign reflecting the nuanced and multifaceted nature of truth, "Truth is ...," on February 26, 2017, during the Oscars (Diaz 2017). CNN hosts now routinely accuse their guests of lying or bending the truth; and chyrons fact-check President Trump as he speaks (Abbruzzese 2016). Headline writers now explicitly call out falsehoods in the stories that follow, rather than leaving it to the text. Journalists readily accuse the President and his staff of indulging in false equivalence, as when he compared Antifa protestors with Nazis and heavily armed white supremacists after the Charlottesville riots.

In parallel, the same outlets have stepped up their already vigorous critiques of the role that technology companies—in particular Facebook, Google, and Twitter—may have played in the election (Cohen 2017; Coren 2017; Entous, Dwoskin and Timberg 2017;

Madrigal 2017; Roose 2017). Stories appear on an almost daily basis highlighting the potential ways in which algorithms and social sharing can combine to spread misinformation (Bajarin 2017; Morgan 2017; Palma and Green 2017). For example, a widely cited BuzzFeed article (Silverman 2016) claimed that the 20 most shared fake news articles Facebook were shared more frequently than the 20 most shared real news articles. Numerous stories have reported on the manipulating of Facebook’s ad system by Russian-affiliated groups, including ads that were targeted at self-described “Jew Haters” (Feldman 2017; Timberg and Dvoskin 2017; Volz 2017). Algorithmic ranking of Google search results has been implicated in amplifying fake reports in the immediate aftermath of the Las Vegas shootings (Machkovech 2017). Lawmakers such as Senator Mark Warner (Dem, VA) have been prominently profiled on account of their outspoken criticism of the Tech industry (Kang 2017). Even Facebook’s own employees have been reported as expressing anxiety over their company’s role in the election (Isaac 2017).

### **A problem with the fake news narrative**

We agree that fake news, and misinformation more generally, are real problems that require scientific attention (Lazer et al. 2017). More broadly, we agree that social media and other web technologies have contributed to more general problems in democratic discourse such as the lack of respect for expert opinion and potentially eroding basis of shared reality (Watts and Rothschild 2017). Nonetheless, we believe that the volume of reporting around fake news, and the role of tech companies in disseminating it, is wildly disproportionate to its likely role in the election outcome, for three reasons.

First, as large as the breathlessly repeated numbers on fake news are made to seem, they are not large when compared with the volume of information to which online users are exposed (Karpf 2017). For example, a New York Times story (Shane and Goel 2017) reported that Facebook had identified over 3,000 ads purchased by fake accounts traced to Russian sources, which generated over \$100,000 in advertising revenue. But Facebook’s advertising revenue in Q4 of 2016 was over \$8,000,000,000, which translates to roughly \$88,000,000 per day; that is, the fake ads in total comprised roughly 0.1% of Facebook’s *daily* advertising revenue. Likewise, a 2016 BuzzFeed report (cite) claimed that the top 20 fake news stories on Facebook “generated 8,711,000 shares, reactions, and comments” between August 1 and Election Day. Again, this sounds like a large number until one considers that Facebook had 1,500,000,000 active monthly users in 2016. If each user took only a single action per day (likely an underestimate), then over the 100 days prior to the election, the 20 stories in BuzzFeed’s study would have accounted for only 0.006% of user actions. Even recent claims that the “real” numbers were much higher than initially reported do not hold up to scrutiny. For example, a New York Times story reported on Oct 30 that “Russian agents...disseminated inflammatory posts that reached 126 million users on Facebook, published more than 131,000 messages on Twitter and uploaded over 1,000 videos to Google’s YouTube service” (Isaac and Wakabayashi 2017). Big numbers indeed. Except that several paragraphs later in the same article, the authors concede that over the same period Facebook users were exposed to 11 trillion posts—roughly 87,000 for every fake exposure—while on Twitter the Russian-linked election tweet represented less than 0.75% of all election-related tweets. On YouTube, meanwhile, the

total number of views of fake Russian videos was around 309,000—compared with the 5 Billion YouTube videos that are watched every day (Donchev 2017).

Second, given what is known about the impact of online information on opinions, even the high-end estimates of fake news penetration would be unlikely to have had a meaningful impact on voter behavior. For example, a recent study by Allcott and Gentzkow (2017) estimates that “the average US adult read and remembered on the order of one or perhaps several fake news articles during the election period, with higher exposure to pro-Trump articles than pro-Clinton articles.” In turn, they estimate that this level of exposure would have changed vote shares by “an amount on the order of hundredths of a percentage point.” As the authors acknowledge, the fake news stories could have been concentrated on specific segments of the population who in turn could have had a disproportionate impact on the election outcome. Also, potentially, individual stories can exert outsized influence over readers’ opinions if they resonate with deeply held beliefs or fears. For all these reasons, fake news and ads could have influenced the election more than their relatively tiny average prevalence suggests. Nevertheless, it would have had to be much larger—at least 100 times larger—to account for Trump’s margin of victory in the key states on which the election outcome depended.

Third, there are other reasons to think that the effect could have been even smaller. Many of the shares are likely to have been made by automated “bots” rather than by humans (Bessi and Ferrara 2016). Even among human users, many impressions that are registered by social media platforms do not correspond to actual consumption. Finally, the sheer outrageousness of the

most popular stories—Pope Francis endorsing Trump; Democrats planning to impose Islamic law in Florida; Trump supporters chanting “We hate Muslims, we hate blacks;” and so on (Ritchie 2016)—made them especially unlikely to have altered voters’ pre-existing opinions of the candidates. Notwithstanding the occasional rube who actually believed that Hillary Clinton was running a pedophilia sex ring out of a Washington pizzeria, such stories were most likely consumed by readers who already agreed with their overall sentiment (Hillary is evil, Trump supporters are racists, etc.) and who shared them either to signal their “tribal allegiance” or simply for entertainment value (“cheer leading”), not because they had been persuaded by the stories themselves, or even necessarily believed them to be accurate.

#### **Turning attention to the “real” news**

To the extent that one considers the 2016 presidential election outcome a failure of democratic decision making, fake news is therefore unlikely to be the main culprit. But if not, then what is? A potentially more serious threat is what Benkler et al (2017a) refer to as “a network of mutually-reinforcing hyper-partisan sites that revive what Richard Hofstadter called ‘the paranoid style in American politics,’ combining decontextualized truths, repeated falsehoods, and leaps of logic to create a fundamentally misleading view of the world.” Unlike the fake news numbers above, engagement with sites like Breitbart News, Infor Wars, and the Daily Caller are substantial—especially in the realm of social media. Nevertheless, the more detailed accounting contained in Benkler et al (Benkler et al. 2017b) shows that by any reasonable metric—including Facebook or Twitter shares, but also referrals from other media sites, number of published stories, etc.—the media ecosystem remains

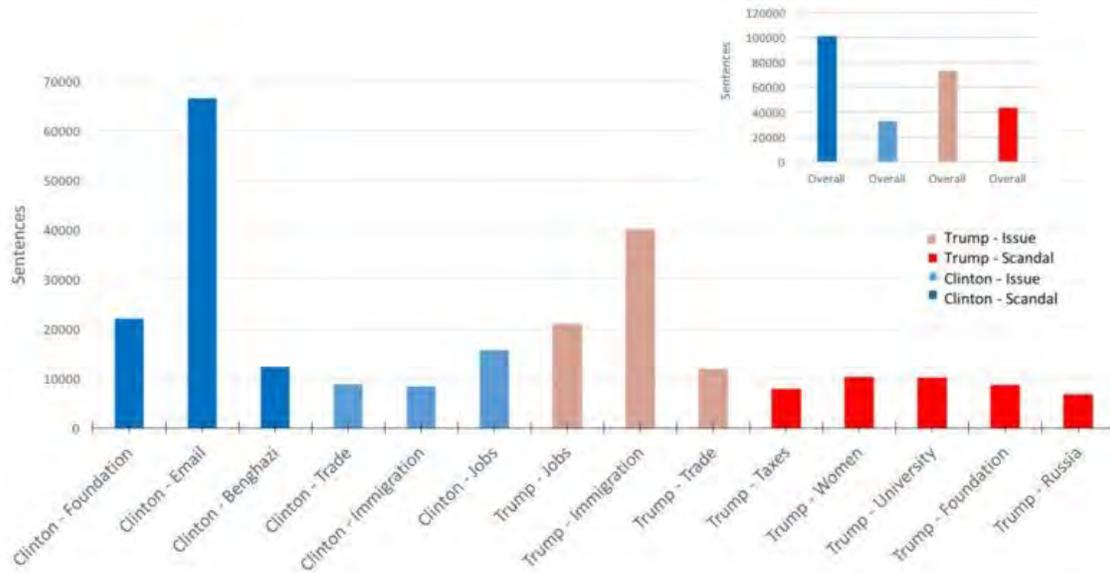
dominated by conventional (and mostly left-of-center) sources such as the Washington Post, the New York Times, Huffington Post, and CNN. In terms of Facebook shares, far and away the largest players were (in order) The New York Times, CNN, Breitbart, Huffington Post, The Hill, and the Washington Post. In terms of Twitter shares, the same five again dominate, albeit in slightly different order (Breitbart drops to fourth after CNN, The New York Times, and The Hill), while in terms of media links Breitbart and Fox News are well down the rankings behind The Washington Post, The New York Times, CNN, and Politico.

Given the attention that these very same news outlets have lavished, post-election, on fake news shared via social media, it may come as a surprise that they themselves dominated social media traffic; but once again some simple math shows why it should not be. While it may have been the case that the twenty most shared fake news stories narrowly outperformed the twenty most shared mainstream stories, the overall volume of stories produced by major newsrooms vastly outnumbers fake news. According to Benkler et al (2017b) “The Washington Post produced more than 50,000 stories over the 18-month period, while the New York Times, CNN, and Huffington Post each published more than 30,000 stories.” (p.28). Presumably not all these stories were about the election, but each such story was also likely reported by many news outlets simultaneously. A rough

estimate of thousands of election-related stories published by the mainstream media is therefore not unreasonable.

What did all these stories talk about? Benkler et al. (2017b) investigated this question also, counting sentences that appeared in mainstream media sources that they then classified as detailing one of several Clinton and Trump related issues. In particular, they classified each sentence as describing either a scandal (e.g. Clinton Emails, Trump Taxes) or a policy issue (Clinton Jobs, Trump Immigration). As shown in Figure 1, they found roughly four times as many Clinton-related sentences that described scandals as policies, whereas Trump-related sentences were one-and-a-half times as likely to be about policy as scandal. Given the sheer number of scandals in which Trump was implicated—sexual assault, the Trump Foundation, Trump University, redlining in his real-estate developments, insulting a gold star family, numerous instances of racist, misogynist, and otherwise offensive speech—and his repeated claims to have been treated unfairly by the media, it is striking that the media devoted more attention to his policies than to his personal failings. Even more striking, one single Clinton-related scandal—her use of a private email server while serving as Secretary of State—accounted for more sentences than all of Trump’s scandals combined (65,000 vs. 40,000) and more than twice as many as were devoted to all of her policy positions.

Figure 1. Count of sentences that appeared in mainstream media sources (e.g. New York Times, Washington Post, Huffington Post, CNN) that were classified as describing either a scandal or a policy issue related to either Trump or Clinton. Reprinted from (Benkler et al. 2017a)



To reiterate, these 65,000 sentences were written not by Russian Hackers, but by professional journalists employed at mainstream news organizations, such as the New York Times, The Washington Post, and the Wall Street Journal.<sup>i</sup> To the extent that voters mistrusted Hillary Clinton, or considered her conduct as Secretary of State to have been negligent and potentially criminal, or were generally unaware of what her policies contained or how they may have differed from Donald Trump’s, these numbers suggest that their views were entirely consistent with the content being produced by mainstream news sources. Moreover, in light of the often bizarre focus of the fake news stories, and given the issues that appear to have energized voters on both sides—the repeal and replace of Obamacare, Hillary Clinton’s email scandal, Donald Trump’s misogyny and racism—it is not far-fetched to speculate that the real news was far more influential in driving the election outcome than the fake news.

**NYT Campaign Coverage**

To shed more light on this possibility, we conducted a more in-depth analysis of a single media source—namely the New York Times.<sup>ii</sup> We gathered two datasets that captured the New York Times’ coverage of the final stage of 2016 presidential election. The first dataset comprised all articles that appeared on the front page of the printed newspaper (399 total) over the last 69 days of the campaign, beginning on Sep 1, 2016 and ending on Nov 8 (election day). The second dataset comprised the full corpus of all 13,481 articles published online over the same period. In both datasets, we first identified all articles that were relevant to the election campaign. We then further categorized each of these articles as belonging to one of three top-level categories:<sup>iii</sup> *Campaign Miscellaneous*, *Personal/Scandal*, and *Policy*. Within *Personal/Scandal* we then further classified the article as focused on Clinton or Trump, and within *Policy* classified it as one of the following: *Policy no details*, *Policy Clinton*

*details, Policy Trump details, and Policy both details:*

*Campaign Miscellaneous* articles focused on the “horse race” elements of the campaign, such as the overall likelihood of victory of the candidates, details of intra-party conflicts, or the mobilization of specific demographic groups (e.g. African Americans, Hispanics). For example, an October 12, story with the headline “Republican Split Over Trump Puts States into Play” described how Clinton’s campaign was taking advantage of Trump’s battle with the Republican party. This article was manifestly about the campaign, but treated it mostly as contest in which a dramatic twist had just taken place. It contained little information that would have helped potential voters understand the candidates’ policy positions and hence their respective agendas as President.

*Personal/Scandal* articles focused on the controversial actions and/or statements of the candidates, either during the election itself or prior to it, as well as on the fallout generated by the controversies. An example of the former would be an Oct 8 article “Tape Reveals Trump Boast About Groping Women,” which discussed the infamous Access Hollywood Tape. An example of the latter would be an October 29 article, “New Emails Jolt Clinton Campaign in Race’s Last Days,” which discussed the impact of the reopening of FBI investigation into Clinton’s private email server on the campaign but provided no new details on the scandal itself. In addition, we classified each Personal/Scandal article as being primarily about Clinton (e.g. emails, Benghazi, the Clinton Foundation) or Trump (e.g. sexual harassment, Trump University, Trump Foundation, etc.).

*Policy* articles mentioned policy issues such as healthcare, immigration, taxation, abortion, or education. Articles coded as *Policy No Details* mentioned policy issues as impacting the campaign, but did not describe the actual policies of the candidates. For example, an Oct 26 article, “Growing Costs of Health Law Pose a Late Test” described Donald Trump attacking Hillary Clinton over the issue of premium increases, but did not mention the policy proposals of the two candidates, nor did it note that due to subsidies the hikes would not affect the actual price paid by 86% of people in marketplaces. *Policy Clinton Details* or *Policy Trump Details* counted articles that mentioned specifics of the Clinton or Trump platforms respectively but not both, while “Policy Both Details” compared the specifics of the two candidates’ platforms. For example, an October 3 article, “Next President Likely To Shape Health Law Fate,” which we coded as *Policy Both Details*, noted that Clinton had endorsed “a new government-sponsored health plan, the so-called public option, to give consumers an additional choice.” It also noted that “Donald J. Trump and Republicans in Congress would go in the direction of less government, reducing federal regulation and requirements so insurance would cost less and no-frills options could proliferate. Mr. Trump would, for example, encourage greater use of health savings accounts, allow insurance policies to be purchased across state lines and let people take tax deductions for insurance premium payments.”

Table 1 shows the results of our analysis. Of the 150 front-page articles that discussed the campaign in some way, we classified slightly over half (79) as Campaign Miscellaneous. Slightly over a third (55) were Personal/scandal with 30 focused on Trump and 25 on Clinton. Finally, just over

10% (16) of articles discussed Policy, of which six had no details, four provided details on Trump policy only, one on Clinton policy only, and five made some comparison between the two candidates' policies. Table 2 shows the corresponding results for the full corpus (see Appendix for details of methods). Of the 1,433 articles that mentioned Trump or Clinton, 291 were devoted to scandals or other personal

matters, while only 70 mentioned policy. Of those only 60 mentioned any details of either candidate's positions. In other words, comparing the two datasets, the number of Personal/Scandal stories for every Policy story ranged from 3.4 (for front page stories) to 4.2. Further restricting to policy stories that contained some detail about at least one candidate's positions, these ratios rise to 5.5 and 4.85 respectively.

Table 1: Front page New York Times stories classified as about the campaign that focused primarily on personal scandals, policy, and miscellaneous matters (see text for coding).

	<b>Campaign n Misc.</b>	<b>Personal/ Scandal Clinton</b>	<b>Personal/ Scandal Trump</b>	<b>Policy No Detail</b>	<b>Policy Clinton Detail</b>	<b>Policy Trump Detail</b>	<b>Policy Both Detail</b>
<b>Sep.</b>	<b>30</b>	<b>6.5</b>	<b>8.5</b>	<b>2</b>	<b>4</b>	<b>1</b>	<b>3</b>
<b>Oct.</b>	<b>33</b>	<b>13.5</b>	<b>17.5</b>	<b>4</b>	<b>0</b>	<b>0</b>	<b>2</b>
<b>Nov.</b>	<b>16</b>	<b>5</b>	<b>4</b>	<b>0</b>	<b>0</b>	<b>0</b>	<b>0</b>
<b>Tot.</b>	<b>79</b>	<b>25</b>	<b>30</b>	<b>6</b>	<b>4</b>	<b>1</b>	<b>5</b>

Table 2: New York Times stories classified in the same manner as above but drawn from the full corpus published at nytimes.com.

	<b>Mention Trump/Cl inton</b>	<b>Persona l/ Scandal Clinton</b>	<b>Personal/ Scandal Trump</b>	<b>Policy No Detail</b>	<b>Policy Clinton Detail</b>	<b>Policy Trump Detail</b>	<b>Policy Both Detail</b>
<b>Total</b>	<b>1,433</b>	<b>87</b>	<b>204</b>	<b>10</b>	<b>13</b>	<b>30</b>	<b>17</b>

As has become clear since the election, there were in fact profound differences between the two candidates' policy positions, and these differences are already proving enormously consequential to the American people. Under President Trump, the Affordable Care Act is being actively undermined, environmental and consumer protections are being rolled back, international alliances and treaties are being threatened, and immigration policy has been thrown into turmoil, among other dramatic changes. If ever there was an election in

which the news media had an obligation to inform citizens of the substantive policy implications of their choice, the election of 2016 should have been such an election. And yet only five out of 150 front page articles that the New York Times ran over the last, most critical months of the election even attempted to make these comparisons evident to their readers, while only ten even mentioned policy details of either candidate.<sup>iv</sup>

*Clinton Email Scandal.* In this context, ten is an interesting number because it is also the number of front-page stories that the New York Times ran on the Hillary Clinton email scandal in just six days, from Oct 29 (the day after FBI director James Comey announced his decision to reopen his investigation of possible wrongdoing by Clinton) through Nov 3, just five days before the election (see Fig 2). In retrospect, one cannot help but be struck by the sheer extent of coverage, day after day, of this one issue on the front page of the nation's leading newspaper. The level of coverage

becomes truly astounding when juxtaposed with the coverage of all policy issues combined. To reiterate, *in just six days the New York Times ran as many cover stories about Hilary Clinton's emails as they did about all policy issues combined in the 69 days leading up to the election.*<sup>v</sup> Nor can this intense focus on the email scandal be written off as inconsequential: the Comey incident and its subsequent impact on Clinton's approval rating among undecided voters was very likely sufficient to have tipped the election (Silver 2017a).

Figure 2: Ten articles on the front page of the New York Times in a six-day period (October 29 to November 3, 2016), discussing the FBI investigation into Secretary Clinton's use of a private email server. In the same time-period there were six front page articles on the dynamics of the campaign, one piece on Trump's business, and zero on public policy of candidates.



*Coverage of Obamacare.* Setting scandals aside for a moment, arguably the most salient policy issue of the election campaign was the success or failure of Obamacare. In the months after Donald Trump's inauguration, The Upshot, the data-centric sub-section of the New York Times published two pieces on Obamacare. The first, "One-Third Don't Know Obamacare and Affordable Care Act Are the Same," published on Feb 7, described some important misconceptions about Obamacare held by large percentages of the American public—for example, that almost 40% (and 47% of Republicans) did not know that repealing Obamacare would cause people to lose Medicaid coverage or subsidies for private insurance. The second, "No, Obamacare isn't in a 'Death Spiral,'" published on March 15, 2017, provided readers with some important details about how Obamacare works. For example, it noted that "because of how subsidies work, people were generally shielded from this year's higher prices." It also noted that while prices had gone up recently, they "were lower than expected in the first few years of the program." The article then went on to make a clear picture of an insurance market that could certainly use improvement, but that the "Obamacare markets will remain stable over the long run, if there are no significant changes."

Given the importance of health care to a large percentage of Americans, including

Trump voters, one might have expected to see detailed explanations of how Obamacare worked and how well it was working in the months leading up to the election. In contrast, the Times' pre-election coverage of Obamacare was both surprisingly sparse (we counted only four front page stories between Sept 1 and Nov 8) and surprisingly negative. As Figure 3 shows, the first article, on October 3, strikes a much gloomier note than the Upshot articles above, stating "Mr. Obama's signature domestic achievement will almost certainly have to change to survive." But then, in a three-day period from Oct 25-27 the Times' tone was even more negative: "Choices fall in health law as costs rise" blares the October 25 headline; "Growing costs of health law pose a late test;" and finally "Many prefer tax penalties to health law." All four articles emphasized troubles in the insurance market, failing to mention that most policy holders have subsidized capped prices (so are insulated from premium hikes), or that the government was spending less than anticipated, or that premiums were rising slower than before Obamacare. Further, none of the articles mention the Medicaid expansion, which has proven to be one of the most popular parts of the bill. Instead, the articles draw their negative inferences from the small fraction of Obamacare users who paid full price in the markets, while ignoring all the features that have made it so popular postelection.

Figure 3. All four front-page articles that touch on the Affordable Care Act (aka ObamaCare), published on (from left to right) October 3, 25, 26, 27.



## Conclusion

Consistent with Benkler et al.'s (2017b) sentence-level count, our analysis of the New York Times' coverage finds that it mostly focused on "dramatic" issues like the horserace or personal scandals, not on the kind of substantive policy issues that would have allowed readers to make informed decisions about the candidates' platforms.<sup>vi</sup> If voters had wanted to educate themselves on issues such as healthcare, immigration, taxes, and economic policy—or how these issues would likely be affected by the election of one or other of the candidates as President—they would not have learned much from reading the New York Times. What they would have learned about was that both candidates were plagued by scandal: Hillary Clinton over her use of a private email server for government business while Secretary of State and also allegations of possible conflicts of interest in the Clinton Foundation; and Trump over his failure to release his tax returns, his past business dealings, Trump University, the Trump Foundation, accusations of sexual harassment and assault, and numerous misogynistic, racist, and otherwise insensitive or offensive remarks.<sup>vii</sup> What

they would also have learned about was the ever-fluctuating state of the horse race: who was up and who was down; who might turn out and who might not; and who was happy or unhappy with whom about what.

To reiterate, we chose to analyze the New York Times' coverage not because we felt it was worse than any other mainstream news organization (again our results are broadly consistent with Benkler et al. (2017b), but rather because it illustrates a broader failure of mainstream journalism to inform their audiences of the very real and very consequential issues at stake. In retrospect, it seems clear that the Times—along with essentially all its peer organizations—made the mistake of assuming that a Clinton victory was inevitable, and were setting themselves as credible critics of the next administration (Silver 2017a). Possibly this mistake arose from the failure of journalists to get of their "hermetic bubble" as some have asserted (Spayd 2017). Possibly it was their misinterpretation of available polling data, which showed all along that a Trump victory, albeit unlikely, was far from inconceivable (Silver 2017b; Watts 2017). These were understandable mistakes, but

they were still mistakes. Yet rather than acknowledging the possible impact that their collective failure of imagination could have had on the election outcome, the mainstream news community has instead focused its critical attention on fake news, Russian hackers, technology companies, algorithmic ranking, the alt-right media, even on “Americans’ inability or unwillingness to sift fact from fiction” (Parker 2017)—that is everywhere but on themselves.

To be fair, journalists were not the only community to be surprised by the outcome of the 2016 election—a great many informed observers, possibly including the candidate himself, failed to take the prospect of a Trump victory seriously. Also to be fair, the difficulty of adequately covering a campaign in which the “rules of the game” were repeatedly upended must surely have been formidable (Hepworth et al. 2016). That said, one could equally argue that Facebook could not have been expected to anticipate the misuse of its advertising platform to seed fake news stories. And one could just as easily argue that the difficulties facing tech companies in trading off between complicity in spreading intentional misinformation on the one hand, and censorship on the other hand, are every bit as formidable as those facing journalists trying to cover Trump. For journalists to excoriate the tech companies for their missteps while barely acknowledging their own reveals an important blind spot in the journalistic profession’s conception of itself.

We have no doubt that journalists take their mission to provide their readers with the information they need “in order to make decisions regarding their lives, and their local and national communities... in an accurate, comprehensive, timely and understandable manner”<sup>viii</sup> seriously. We note, however, that this mission implicitly

assumes that journalists are passive observers of events rather than active participants whose choices about what to cover and how to cover it do not meaningfully influence the events in question. Given the disruption visited upon the print news business model since the beginning of the 21<sup>st</sup> century, journalists can perhaps be forgiven for seeing themselves as helpless bystanders in an information ecosystem that is increasingly centered around social media and search. But even if the news media has ceded much of its distribution power to technology companies, its longstanding ability to “set the agenda” (McCombs and Shaw 1972)—that is, to determine what counts as news to begin with—remains formidable. In sheer numerical terms, the information to which voters were exposed during the election campaign was overwhelmingly produced not by fake news sites or even by alt-right media sources, but by household names like the New York Times, the Washington Post, and CNN. Without discounting the role played by malicious Russian hackers and naïve tech executives, we believe that fixing the information ecosystem is much more about improving the real news than stopping the fake stuff.

*The authors are grateful to William Cai for valuable research assistance.*

#### **Appendix: Analysis of the Full Corpus**

What the NYT puts on its front page of its print edition is important, but not necessarily representative of how many readers encounter the news, either because they navigate directly to individual articles from social media sources (mostly Facebook and Twitter), or because articles at nytimes.com can appear in different places at different times. To check that our conclusions regarding coverage of the campaign were not badly biased by restricting our sample to

the front page only, we also coded the entire corpus of all articles published on nytimes.com during the same period.

Because this sample is much larger (13,481 vs. 399) we coded them using a combination of machine classification and hand coding.

First, we scraped the headline and first paragraph, if given by the API, of all NYT articles from 9/1/2016 to 11/9/2016 using the archive API for all articles that included the words Clinton or Trump. Note: this criterion that would include all campaign-related articles but might also include potentially non-campaign related articles (e.g. about Bill Clinton or Ivanka Trump). Next, we compiled a list of words (details below) delineating three categories of article: *campaign* (focused on horse race and how people react to events); *policy* (focused on a policy issue); and *personal/scandal*. For each article, we check if any of the words in the article begin with one of the stems in our word list. For example, if an article contains the word 'immigration', we notice that it starts with 'immigrant', which is one of our policy words, and we mark it as policy article. Then, for all articles which are marked as candidate policy articles, we hand-code them into the four sub-categories and toss articles into campaign miscellaneous if they do not actually cover any policy.

Finally, we hand-coded the Policy articles as Policy No Details, Policy Clinton Details, Policy Trump Details, or Policy Both Details using the same criteria as above.

Word list for Clinton/Trump Categories:

1) Clinton personal/scandal words: ['email', 'benghazi', 'foundation', 'road']

2) Trump personal/scandal words: ['russia', 'foundation', 'university', 'woman', 'women', 'tax', 'sexual assault', 'golf', 'tape', 'kiss']

3) Policy words for both candidates are taken from the list of issues covered by On the Issues: ['abortion', 'budget', 'civil rights', 'corporation', 'crime', 'drug', 'education', 'energy', 'oil', 'environment', 'family', 'families', 'children', 'foreign policy', 'trade', 'reform', 'government', 'gun', 'health', 'security', 'immigra', 'technology', 'job', 'principl', 'value', 'social security', 'war', 'peace', 'welfare', 'poverty', 'econom', 'immigrat', 'immigran']

4) Campaign words for both candidates: ['fundraise', 'ads', 'advertisements', 'campaign', 'trail', 'rally', 'endorse', 'outreach', 'ballot', 'vote', 'electoral', 'poll', 'donat', 'turnout', 'margin', 'swing state']

## References

- Abbruzzese, Jason. 2016. "CNN Rolled Out a Pretty Bold Way to Fact-Check Donald Trump." in Mashable.
- Albright, Jonathan. 2017. "Cambridge Analytica: the Geotargeting and Emotional Data Mining Scripts." in *Medium*.
- Allcott, Hunt, and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." National Bureau of Economic Research.
- Bajarin, Tim. 2017. "Forget Politics: Why 'Fake News' Should Infuriate You." in *PC Magazine*.
- Benkler, Yochai, Robert Faris, Hal Roberts, and Ethan Zuckerman. 2017a. "Study: Breitbart-led right-wing media ecosystem altered broader

- media agenda." in *Columbia Journalism Review*.
- Benkler, Yochai, Hal Roberts, Robert Faris, Bruce Etling, Ethan Zuckerman, and Nikki Bourassa. 2017b. "Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election." Berkman Klein Center for Internet & Society Research Paper.
- Bessi, Alessandro, and Emilio Ferrara. 2016. "Social bots distort the 2016 US Presidential election online discussion."
- Cohen, Noam. 2017. "Silicon Valley is Not Your Friend." in *New York Times*.
- Coren, Michael. 2017. "Ending fake news means changing how Wall Street values Facebook and Twitter." in *Quartz*.
- Delaney, Kevin J. 2017. "Filter bubbles are a serious problem with news, says Bill Gates." in *Quartz*. 21 February 2017.
- Diaz, Ann-Christine. 2017. "'The Truth is Hard,' Says The New York Times' First-Ever Oscars Ad." in *AdAge*.
- Donchev, Danny. 2017. "36 Mind Blowing YouTube Facts, Figures, and Statistics - 2017." in *Fortune Lords*.
- El-Bermaway, Mostafa. 2016. "Your Filter Bubble is Destroying Democracy." in *Wired*.
- Entous, Adam, Elizabeth Dwoskin, and Craig Timberg. 2017. "Obama tried to give Zuckerberg a wake-up call over fake news on Facebook." in *Washington Post*.
- Fahri, Paul. 2017. "The Washington Post's New Slogan Turns Out to be an Old Saying." in *Washington Post*.
- Feldman, Brian. 2017. "Facebook Hands Over Data to Congress and Promises Better Disclosure on Political Ads." in *New York Magazine*.
- Hepworth, Shelley, Vanessa Gezari, Kyle Pope, Cory Schouten, Carlett Spike, David Uberti, and Pete Vernon. 2016. "Covering Trump: An oral history of an unforgettable campaign." in *Columbia Journalism Review*.
- Isaac, Mike. 2017. "At Facebook, Hand-Wringing Over a Fizz for Fake Content." in *New York Times*.
- Isaac, Mike, and Daisuke Wakabayashi. 2017. "Russian Influence Reached 126 Million Through Facebook Alone." in *New York Times*.
- Kang, Cecilia. 2017. "Mark Warner: Tech Millionaire Who Became Tech's Critic in Congress." in *New York Times*.
- Karpf, David. 2017. "People are hyperventilating over a study of Russian propaganda on Facebook. Just breathe deeply." in *Washington Post*.
- Lazer, David, Matthew Baum, Yochai Benkler, Adam Berinsky, Kelly Greenhill, Filippo Menczer, Miriam Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A

- Sloman, Cass R Sunstein, Emily Thorson, Duncan J. Watts, and Jonathan Zittrain. 2017. "The science of fake news." Under review.
- Machkovech, Sam. 2017. "Google admits citing 4chan to spread fake Vegas shooter news." in *Ars Technica*.
- Madrigal, Alexis C. 2017. "12 October 2017." in *The Atlantic*.
- McCombs, Maxwell E, and Donald L Shaw. 1972. "The agenda-setting function of mass media." *Public Opinion Quarterly* 36(2):176-87.
- Morgan, Jonathon. 2017. "Facebook and Google need to own their role in spreading misinformation -- and fix it." in *CNN*.
- Palma, Bethania, and Vinny Green. 2017. "Misinfluencers, Inc.: How Fake News Is Reaching Millions Using Verified Facebook Accounts." in *Snopes*.
- Parker, Emily. 2017. "Silicon Valley Can't Destroy Democracy Without Our Help." in *New York Times*.
- Rees, Annie. 2016. "In NYT's Hillary Clinton Coverage, An Obsession With 'Clouds' And 'Shadows'." in *Talking Points Memo*.
- Ritchie, Hannah. 2016. "Read all about it: The biggest fake news stories of 2016." in *CNBC*.
- Roose, Kevin. 2017. "We Asked Facebook 12 Questions About the Election, and Got 5 Answers." in *New York Times*.
- Rutenberg, Jim. 2017. "RT, Sputnik and Russia's New Theory of War." in *New York Times Magazine*.
- Shane, Scott, and Vindu Goel. 2017. "Fake Russian Facebook Accounts Bought \$100,000 in Political Ads." in *New York Times*.
- Silver, Nate. 2017a. "The Comey Letter Probably Cost Clinton The Election." in *FiveThirtyEight*.
- . 2017b. "The Media Has A Probability Problem." in *FiveThirtyEight*.
- Silverman, Craig. 2016. "This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook." in *BuzzFeed*.
- Spayd, Liz. 2017. "Seeking More Voices, Even if Some Don't Want to Hear Them." in *New York Times*.
- Timberg, Craig. 2017. "Russian propaganda may have been shared hundreds of millions of times, new research says." in *Washington Post*.
- Timberg, Craig, and Elizabeth Dwoskin. 2017. "Facebook takes down data and thousands of posts, obscuring reach of Russian disinformation." in *Washington Post*.
- Volz, Dustin. 2017. "Facebook will help investigators release Russia ads, Sandberg tells Axios." in *Reuters*.
- Watts, Duncan J, and David Rothschild. 2017. "Rebuilding legitimacy in a post-truth age." *Medium* Jan 17.

Watts, Duncan J. 2017. "The non-inevitability of Donald Trump (and almost everything)." in *Medium*.

---

<sup>i</sup> Breitbart and other "alt-right" media organizations (e.g. TruthFeed, the Daily Caller) were included in this sample, but a separate count in which the two largest sites (Breitbart and the Daily Caller) were removed yielded qualitatively similar results (personal correspondence with Y. Benkler).

<sup>ii</sup> We chose the Times for two reasons: first, because its broad reach both among policy elites and ordinary citizens means that the Times has singular influence on public debates; and second, because its reputation for serious journalism implies that if the Times did not inform its readers of the issues then it is unlikely that such information was widely available anywhere.

<sup>iii</sup> Every front-page article was read by two researchers and coded for these three topline categories and sub-categories, using only the text that appeared on the actual front-page (not on what may be continued on future pages). There was very little disagreement between the two researchers. Most critically there was agreement over the articles that covered policies of both candidates and just one article difference on the total number of articles that included policy. For simplicity, the authors reviewed all disagreements together, by hand, and we report from that dataset.

<sup>iv</sup> Perhaps even more damning, even the tiny minority of articles that did attempt to deal with policy details skimmed over some important ones. For example, the Oct 3 article from earlier did not mention the key provision of cutting the Medicaid expansion, nor did it mention any consequences of eliminating Essential Benefits (e.g. making healthcare insurance unaffordable for people with pre-existing conditions). Given that healthcare was arguably the most important salient issue of the election to many voters—these omissions seem significant.

<sup>v</sup> If one counts the two additional front page articles on Clinton's email that appeared on November 6 and November 7 (the latter describing Comey's notification to Congress that the renewed investigation had uncovered nothing new), then these numbers look even more lopsided.

<sup>vi</sup> One of the authors, David Rothschild, also covered the horse race, writing a piece in the New York Times on how the polling margin of error is larger than advertised. He was featured in two other articles, one where he was an outlier in showing Trump to win Florida and another where he argued that the winner of Pennsylvania will win the election.

<sup>vii</sup> Even when covering scandals, the coverage was less about what happened than the perception of what happened, as evidenced by the frequent references to "shadows," "clouds," and other implied rather than actual wrong-doing. (Rees 2016).

<sup>viii</sup> From the website of the Society of Professional Journalists

## Personalized Information Environments and Their Potential Consequences for Disinformation

Deen Freelon, University of North Carolina

Ever since the internet started to permeate everyday life, academics and journalists have fretted about the ill effects of personalized information environments. This idea was first popularized by Negroponte (1996) in the mid-1990s as the “Daily Me,” and developed into its best-known form primarily by Sunstein (2007) and Pariser (2011). Other closely-related concepts include echo chambers, cyberbalkanization, filter bubbles, and selective exposure to media content. Personalized information environments are digital content delivery systems configured to suit the idiosyncratic tastes of a single user. They may be constructed manually, as by following particular accounts on a social media platform; or automatically, as when a recommender system serves content on the basis of a user’s prior behaviors. Many contemporary sociotechnical systems, including Facebook and Twitter, incorporate both manual and automated components to generate each user’s individual feed.

These environments can become problematic when considered from normative democratic perspectives that values civility and the exchange of information between people and groups that disagree fundamentally. Perhaps the best-known of these in political science and communication circles is deliberative democracy, which upholds cross-cutting interaction as a core priority (Carpini, Cook, & Jacobs, 2004; Freelon, 2010; Mutz, 2006). This perspective condemns as undemocratic media tools and environments that allow people to read only content that flatters their

political preconceptions while screening out dissenting opinions. The fear is that people will use such tools as much as possible, polarizing public opinion and perhaps even rendering civil debate between two opposing sides impossible (Gutmann & Thompson, 2004; Schkade, Sunstein, & Hastie, 2010; Sunstein, 2007).

The deliberative objection can be unpacked into two falsifiable research questions:

To what extent do people with the technical capacity to create personalized information environments use them to create ideologically airtight “echo chambers”?

To what extent do people who use highly polarized information environments exhibit extreme political opinions and intolerance of opposing viewpoints?

Given the topic of this conference, we might also consider a third research question:

How is the use of personalized information environments related to use of and trust in disinformation content?

The remainder of this paper will briefly review the empirical literature addressing these questions and conclude by raising research questions of particular relevance for disinformation research.

### **Do echo chambers exist?**

The foundational theoretical work in this area made several key assumptions, among them 1) that it is possible to construct hermetically-sealed information bubbles through which no (or almost no) contrary views can penetrate; and 2) that given a choice between such an environment and a more ideologically balanced information diet, most people would prefer the former (Pariser, 2011, pp. 12, 18; Sunstein, 2007, p. 1). Most empirical work on the topic suggests that neither of these assumptions generally hold. Instead, people tend to exhibit moderate to strong preferences for ideologically-consistent content while failing to routinely reject content from the other side (Bakshy, Messing, & Adamic, 2015; Barberá, Jost, Nagler, Tucker, & Bonneau, 2015; Dvir-Gvirsman, Tsfati, & Menchen-Trevino, 2016; Garrett, 2009; Gentzkow & Shapiro, 2011; Hanna et al., 2013; Hargittai, Gallo, & Kane, 2008; Kobayashi & Ikeda, 2009). Some studies have found differences in the degrees of ideological self-segregation between liberals and conservatives (Barberá et al., 2015; Boutyline & Willer, 2017; Hanna et al., 2013) as well as between political issues (Bakshy et al., 2015; Barberá et al., 2015; Himmelboim, McCreery, & Smith, 2013). Overall, the information environments most people choose might look more slanted than Sunstein and Pariser would prefer, but they are less so than predicted.

### **Do personalized information environments polarize?**

We might dismiss the requirement that personalized information environments completely filter out all opinion-inconsistent content as a strawman. It makes more sense to consider the degree of ideological bias: a completely biased environment would fulfill the criteria of an echo chamber, but these are exceedingly rare if they exist at all. That

said, it is still possible that opinion extremity or polarization is correlated with the degree of one-sidedness of one's information environment. The relevant research tends to support this prediction (Garrett et al., 2014; Knobloch-Westerwick, 2012; Levendusky, 2013a, 2013b; Stroud, 2010), although at least one study showed the opposite: that individuals with ideologically diverse social networks hold more polarized feelings toward Democrats and Republicans and identify more strongly as liberal or conservative (Lee, Choi, Kim, & Kim, 2014). That single study notwithstanding, a firm conclusion based on the preponderance of the empirical evidence would be that the more one's media environment looks like an echo chamber, the more extreme one's viewpoints are. There is currently no consensus on the causal direction of this relationship.

### **Methodological challenges**

The original conceptual descriptions of personalized information environments offered specific sites like Google and Reddit as examples, but few people rely on a single service for all the information they consume. If the chief concern is the diversity of people's overall information diets, all possible media types should ideally be considered, whether TV, radio, print, or digital. For obvious reasons, this presents formidable methodological challenges—tracking one person's complete media intake over time is difficult enough, and most researchers would want to track multiple individuals. To my knowledge, no study has yet been able to accomplish this in a rigorous way, although Dvir-Gvirsman et al. (2016) come the closest for online content sources. Consequently, much of the research that purports to examine personalized information environments analyzes only one platform or medium. While a single social media feed may be considered an

information environment unto itself, such research does not directly address the question of overall information diversity on an individual level. Thus, we should not infer that people whose Twitter communication patterns are highly polarized (for example) necessarily experience a media diet that is highly polarized overall. This is not to say that any of the studies cited above do so, but only to point out that some of them define “echo chamber” at the platform level rather than the individual level. Such research is nevertheless still useful, as it indicates whether specific platforms could contribute to individual-level polarization.

### **Possible effects on disinformation**

While misinformation in the sense of unintentional false beliefs has been a long-standing concern of communication research (see Bode & Vraga, 2015; Nyhan, 2010; Weeks, 2015), studies of maliciously-distributed disinformation are much rarer (exceptions include Lewandowsky, Stritzke, Freund, Oberauer, & Krueger, 2013; and Martin, 1982). Studies that analyze the new wave of political disinformation are just now beginning to emerge (Faris et al., 2017; Lazer et al., 2017; Vargo, Guo, & Amazeen, 2017). Thus far, few if any empirical studies have directly addressed the consequences of polarized information environments for disinformation and its effects. However, the following topics would be worth exploring given what we already know:

***Ideological asymmetries in the consumption, distribution, and acceptance of disinformation.*** The following findings support the prediction that conservatives will engage with disinformation substantially more than liberals (and perhaps also centrists):

News fabricators have anecdotally suggested that right-leaning individuals in general, and Trump supporters in particular, are disproportionately susceptible to opinion-reinforcing disinformation online (Dewey, 2016; Sydell, 2016).

An informal analysis by the publication PCWorld found that a Facebook account constructed to appear conservative attracted ten fabricated stories, while one constructed to appear liberal attracted none (Hachman, 2017).

Conservatives have traditionally distrusted mainstream news more than liberals, and this gap has widened in recent years (Barthel & Mitchell, 2017; Swift, 2016). This may drive them toward friendly disinformation sources that lack the perceived bias of traditional news outlets.

### ***Distrust in factual claims generally.***

Conservative distrust for news long predates the current wave of disinformation. But disinformation itself may drive people to adopt distrust as a default position for novel claims about the world, or at least to disbelieve them more than otherwise. Consider:

Corporate purveyors of disinformation, including the tobacco and oil industries, successfully sowed public doubt about political issues affecting their profits for decades (Oreskes & Conway, 2010).

A 2017 BuzzFeed/Ipsos poll found that 54% of respondents distrust news on Facebook, in part because it has failed to stop disinformation from spreading (Silverman, 2017).

In July 2017, only 45% of Trump voters believed Donald Trump Jr. met with a Russian lawyer to seek damaging information on Hillary Clinton before the 2016 election, even though he publicly admitted to doing so (Durkheimer, 2017).

***Misuse and abuse of the rhetoric of “critical thinking.”*** Many purveyors and consumers of disinformation claim to be critical thinkers. This may reduce the efficacy of critical thinking-based solutions to disinformation, especially in polarized information environments.

Some of the best-known proponents of the flat-earth misbelief claim to “look at things objectively with a critical eye” (Ambrose, 2017).

A January 2010 print ad for the media outlet Russia Today compared the feasibility of anthropogenic climate change to that of space aliens. The ad’s fine print encouraged readers to “question more” and “challeng[e] the accepted view” (“Russia Today,” 2010).

Some teens consider their high level of trust in the information quality of Google’s top search results as a form of critical thinking (boyd, 2017).

## References

- Ambrose, G. (2017, July 7). These Coloradans say Earth is flat. And gravity’s a hoax. Now, they’re being persecuted. Retrieved September 29, 2017, from <http://www.denverpost.com/2017/07/07/colorado-earth-flat-gravity-hoax/>
- Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, 348(6239), 1130–1132.
- Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological Science*, 26(10), 1531–1542.
- Barthel, M., & Mitchell, A. (2017, May 10). Americans’ Attitudes About the News Media Deeply Divided Along Partisan Lines. Retrieved September 28, 2017, from <http://www.journalism.org/2017/05/10/americans-attitudes-about-the-news-media-deeply-divided-along-partisan-lines/>
- Boutyline, A., & Willer, R. (2017). The social structure of political echo chambers: Variation in ideological homophily in online networks. *Political Psychology*, 38(3), 551–569.
- boyd, danah. (2017, January 5). Did Media Literacy Backfire? Retrieved September 29, 2017, from <https://points.datasociety.net/did-media-literacy-backfire-7418c084d88d>
- Carpini, M. X. D., Cook, F. L., & Jacobs, L. R. (2004). Public deliberation, discursive participation, and citizen engagement: A review of the empirical literature. *Annu. Rev. Polit. Sci.*, 7, 315–344.
- Dewey, C. (2016, November 17). Facebook fake-news writer: ‘I think Donald Trump is in the White House because of me.’ Retrieved September 28, 2017, from

- <https://www.washingtonpost.com/news/the-intersect/wp/2016/11/17/facebook-fake-news-writer-i-think-donald-trump-is-in-the-white-house-because-of-me/>
- Durkheimer, M. (2017, July 27). “Fake News” Leads To No News For The Trump Base. Retrieved September 29, 2017, from <https://www.forbes.com/sites/michaeldurkheimer/2017/07/27/fake-news-leads-to-no-news-for-the-trump-base/>
- Dvir-Gvirsman, S., Tsfati, Y., & Menchen-Trevino, E. (2016). The extent and nature of ideological selective exposure online: Combining survey responses with actual web log data from the 2013 Israeli Elections. *New Media & Society, 18*(5), 857–877. <https://doi.org/10.1177/1461444814549041>
- Faris, R. M., Roberts, H., Etling, B., Bourassa, N., Zuckerman, E., & Benkler, Y. (2017). *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election*. Berkman Klein Center for Internet & Society.
- Freelon, D. (2010). Analyzing online political discussion using three models of democratic communication. *New Media & Society, 12*(7), 1172–1190.
- Garrett, R. K. (2009). Echo chambers online?: Politically motivated selective exposure among Internet news users. *Journal of Computer-Mediated Communication, 14*(2), 265–285.
- Garrett, R. K., Gvirsman, S. D., Johnson, B. K., Tsfati, Y., Neo, R., & Dal, A. (2014). Implications of Pro- and Counterattitudinal Information Exposure for Affective Polarization. *Human Communication Research, 40*(3), 309–332. <https://doi.org/10.1111/hcre.12028>
- Gentzkow, M., & Shapiro, J. M. (2011). Ideological Segregation Online and Offline. *The Quarterly Journal of Economics, 126*(4), 1799–1839. <https://doi.org/10.1093/qje/qjr044>
- Gutmann, A., & Thompson, D. F. (2004). *Why deliberative democracy?* Princeton: Princeton University Press.
- Hachman, M. (2017, September 7). Just how partisan is Facebook’s fake news? We tested it. Retrieved September 28, 2017, from <https://www.pcworld.com/article/3142412/windows/just-how-partisan-is-facebooks-fake-news-we-tested-it.html>
- Hanna, A., Wells, C., Maurer, P., Friedland, L., Shah, D., & Matthes, J. (2013). Partisan Alignments and Political Polarization Online: A Computational Approach to Understanding the French and US Presidential Elections. In *Proceedings of the 2Nd Workshop on Politics, Elections and Data* (pp. 15–22). New York, NY, USA: ACM. <https://doi.org/10.1145/2508436.2508438>
- Hargittai, E., Gallo, J., & Kane, M. (2008). Cross-ideological discussions among conservative and liberal bloggers. *Public Choice, 134*(1), 67–86.

- Himmelboim, I., McCreery, S., & Smith, M. (2013). Birds of a Feather Tweet Together: Integrating Network and Content Analyses to Examine Cross-Ideology Exposure on Twitter. *Journal of Computer-Mediated Communication, 18*(2), 40–60. <https://doi.org/10.1111/jcc4.12001>
- Knobloch-Westerwick, S. (2012). Selective Exposure and Reinforcement of Attitudes and Partisanship Before a Presidential Election. *Journal of Communication, 62*(4), 628–642. <https://doi.org/10.1111/j.1460-2466.2012.01651.x>
- Kobayashi, T., & Ikeda, K. (2009). Selective exposure in political web browsing. *Information, Communication & Society, 12*(6), 929–953.
- Lazer, D., Baum, M., Grinberg, N., Friedland, L., Joseph, K., Hobbs, W., & Mattsson, C. (2017). *Combating Fake News: An Agenda for Research and Action*. Shorenstein Center on Media, Politics, and Public Policy.
- Lee, J. K., Choi, J., Kim, C., & Kim, Y. (2014). Social Media, Network Heterogeneity, and Opinion Polarization. *Journal of Communication, 64*(4), 702–722. <https://doi.org/10.1111/jcom.12077>
- Levendusky, M. (2013a). Partisan Media Exposure and Attitudes Toward the Opposition. *Political Communication, 30*(4), 565–581. <https://doi.org/10.1080/10584609.2012.737435>
- Levendusky, M. (2013b). Why Do Partisan Media Polarize Viewers? *American Journal of Political Science, 57*(3), 611–623. <https://doi.org/10.1111/ajps.12008>
- Lewandowsky, S., Stritzke, W. G., Freund, A. M., Oberauer, K., & Krueger, J. I. (2013). Misinformation, Disinformation, and Violent Conflict: From Iraq and the “war on Terror” to Future Threats to Peace. *American Psychologist, 68*(7), 487–501. <https://doi.org/10.1037/a0034515>
- Martin, L. J. (1982). Disinformation: An instrumentality in the propaganda arsenal. *Political Communication, 2*(1), 47–64. <https://doi.org/10.1080/10584609.1982.9962747>
- Mutz, D. C. (2006). *Hearing the other side: Deliberative versus participatory democracy*. Cambridge University Press. Retrieved from <http://books.google.com/books?hl=en&lr=&id=r2i3cog2IXEC&oi=fnd&pg=PA1&dq=hearing+the+other+side&ots=ydYzxHeEco&sig=pCaHNvoHPP3ucojrMtu4cGyRiNc>
- Negroponte, N. (1996). *Being Digital*. New York, NY: Vintage.
- Nyhan, B. (2010). Why the “Death Panel” Myth Wouldn’t Die: Misinformation in the Health Care Reform Debate. *The Forum, 8*(1). <https://doi.org/10.2202/1540-8884.1354>
- Oreskes, N., & Conway, E. M. (2010). *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming* (1 edition). Bloomsbury Press.

- Pariser, E. (2011). *The filter bubble: What the Internet is hiding from you*. New York: Penguin.
- Russia Today: "Climate change" Print Ad by McCann London. (2010, January). Retrieved September 29, 2017, from <https://www.coloribus.com/adsarchive/prints-outdoor/russia-today-climate-change-14024155/>
- Schkade, D., Sunstein, C. R., & Hastie, R. (2010). When Deliberation Produces Extremism. *Critical Review*, 22(2–3), 227–252. <https://doi.org/10.1080/08913811.2010.508634>
- Silverman, C. (2017, April 13). Most Americans Don't Trust The News They See On Facebook, According To A New Survey. Retrieved September 29, 2017, from <https://www.buzzfeed.com/craigsilverman/americans-cant-agree-on-what-news-is-on-facebook>
- Stroud, N. J. (2010). Polarization and Partisan Selective Exposure. *Journal of Communication*, 60(3), 556–576. <https://doi.org/10.1111/j.1460-2466.2010.01497.x>
- Sunstein, C. (2007). *Republic. com 2.0*. Princeton, NJ: Princeton University Press.
- Swift, A. (2016, September 14). Americans' Trust in Mass Media Sinks to New Low. Retrieved September 28, 2017, from <http://news.gallup.com/poll/195542/americans-trust-mass-media-sinks-new-low.aspx>
- Sydell, L. (2016, November 23). We Tracked Down A Fake-News Creator In The Suburbs. Here's What We Learned. Retrieved September 28, 2017, from <http://www.npr.org/sections/alltechconsidered/2016/11/23/503146770/npr-finds-the-head-of-a-covert-fake-news-operation-in-the-suburbs>
- Vargo, C. J., Guo, L., & Amazeen, M. A. (2017). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. *New Media & Society*, 1461444817712086. <https://doi.org/10.1177/1461444817712086>
- Weeks, B. E. (2015). Emotions, Partisanship, and Misperceptions: How Anger and Anxiety Moderate the Effect of Partisan Bias on Susceptibility to Political Misinformation. *Journal of Communication*, 65(4), 699–719. <https://doi.org/10.1111/jcom.12164>

## Making Sense of Information and Judging its Credibility

**Natalie Jomini Stroud**, University of Texas at Austin

**Emily Thorson**, Syracuse University

**Dannagal Young**, University of Delaware

### Introduction

Our collective charge was to write about how people make sense of information and judge its credibility. While this brief paper does not cover these topics comprehensively; it does provide summaries of three areas in which we believe interventions could be especially effective at reducing the spread and impact of misinformation. In particular, we discuss the roles of media content and structure, social identity, and information processing styles.

### The role of media content and structure

It almost goes without saying that the content of information can affect whether people find it credible. Weak arguments, disreputable sources, poor grammar, and out-of-date formatting are all examples of characteristics that can affect whether people see information as credible. Beyond these predictable instances, however, research has shown a number of other ways in which information can inspire trust, or a lack thereof.

Different types of reporting affect how credible we find news media coverage. For example, discussing politics in terms of winning and losing, as opposed to emphasizing issues, can lead to greater distrust of the media (Cappella & Jamieson, 1997; Hopmann, Shehata, & Strömbäck, 2015). As another example, the presence of analysis and interpretation in news reporting may increase distrust because it offers more points with which an audience can disagree (Jones, 2004). This is even more possible

given that this style of analytic journalism has increased over the past century (Barnhurst & Mutz, 1997). As a final example, the prevalence of critiques of the media's credibility – many as reported by the media! – can undermine trust in the media (Watts, Domke, Fan, & Shah, 1999).

Fact-checking, a specific subset of news reporting, also can affect the ways in which people make sense of what is factual and what is not. Three examples help to illustrate this idea. First, fact-checking that comes from an unexpected source (e.g. a Republican calling something untrue that would normally be endorsed by Republicans) can be more effective at reducing misperceptions (Berinsky, 2015). Second, when the content of the fact-check contains images or text that calls to mind the original misinformation, people are less responsive to the fact-check. In one study attempting to correct the false claim that “Feisal Abdul Rauf, the Imam backing the proposed Islamic cultural center and mosque, is a terrorist-sympathizer,” including images of the Imam in traditional Arab clothing reduced the power of the corrected information at reducing misperceptions (Garrett, Nisbet, & Lynch, 2013). Third, correcting misinformation is more effective when an alternative cause can be described as opposed to merely saying that the original information is incorrect (Nyhan & Reifler, 2015). For instance, when trying to counter the misperception that vaccines cause autism, it would be more effective to explain what does cause autism

than merely stating that there is not an association between the two.

The sheer amount of information also can affect how people make sense of it. When given two pieces of information, one that reflects a person's belief (e.g. a pro-life article and a person is pro-life) and one that articulates a different point of view (e.g. a pro-choice article and a person is pro-life), people are more likely to look at the counter-attitudinal information. When given ten pieces of information, however, people select a higher percentage of pro- than counter-attitudinal information (Fischer, Shulz-Hardt, & Frey, 2008). One reason for this pattern is that when people encounter more information, they decide what to look at based on its perceived quality. And it just so happens that information judged to be of high quality tends to agree with our beliefs. So when people are inundated with information, pro-attitudinal information that is judged as higher quality and is more likely to be selected.

The structures that we put into place also influence the information that we encounter, and help us to tame the deluge of information (Neuman, 2016). Search engines and news aggregators have become increasingly popular places for gaining information. For some, algorithmic news selection is trusted due to "the perceived independence of algorithms, which most considered to be less influenced by political and editorial agendas than journalists" (Kantar Media, 2016, p. 64). New structures can reshape how the public determines what is credible.

As this brief review demonstrates, the information itself can signal trustworthiness and correct misinformation more or less effectively. Because information can be tailored, we propose this as one important

way to think about how people make sense of information and judge its credibility.

### **The role of social identity**

How we see ourselves, as well as how we want others to see us, both profoundly shape what information we choose to read, believe, and share. Even partisan-driven motivated reasoning is in many ways a social phenomenon – we seek out information that affirms our social identity as a partisan and avoid information that undermines it (Bennett 2012; Green, Palmquist, & Schickler, 2004). In an era of social media, the role of social identity in information choice is magnified because so many of these choices are public. Twitter and Facebook provide countless opportunities to signal to our social networks – via liking, retweeting, sharing, or posting – what content we endorse or oppose (Shin & Thorson 2017).

Much existing research on social identity and information consumption has focused on how social cues affect what people read, process, and recall. For example, Messing and Westwood (2014) find that social endorsements are extraordinarily powerful in directing audience attention to a news story – even more so, in fact, than the partisanship of the news source. And when it comes to misinformation and rumors, social signals may carry an outside importance because other cues that people rely on to determine credibility (for example, authoritative sources) are often absent (Metzger, Flanagin, & Medders 2010). However, for the purposes of designing interventions to stop the spread of misinformation, leveraging social identity may be more effective at the point of distribution rather than at the point of reception. Because most information online spreads through relatively few powerful nodes, interventions that discourage those

nodes from sharing false information can have large downstream effects (Goel et al., 2015).

Several factors shape a person's decision about whether or not to publicly share or endorse a piece of information. Some of these factors are relatively stable -- for example, people high in conflict avoidance are unlikely to share content they believe will be controversial (Vraga et al., 2015), but others are more malleable. One area that may be especially ripe for intervention is normative beliefs – specifically, concerns over creating negative impressions. People are less likely to endorse content if they believe that others might disapprove of that endorsement (Marder et al., 2016). Of course, these normative beliefs almost always compete with other motivations. For example, a strong Democrat might simultaneously crave the social reward that comes from sharing an unverified rumor about Donald Trump with her liberal friends while also worrying about the potential backlash if that rumor turns out to be false. Given these competing beliefs, a successful intervention designed to stop the spread of misinformation might take on two tasks. First, it could *increase* the salience of the normative belief that spreading misinformation will incur negative social consequences. Second, it could emphasize the potential social costs of spreading highly partisan information, potentially by building on a growing body of research suggesting that Americans are increasingly turned off by public displays of partisanship (Klar & Krupnikov, 2016).

### **The role of information processing styles**

Psychological factors play a large role in how people process information. In the context of social media, where an abundance of information in a readily accessible, headline-oriented format is sandwiched

between photos of friends' children and videos of panda bears, promoting thoughtful, rational consideration of news content and source credibility is particularly challenging (Ecker et al., 2014). Dual-process models of judgment formation posit that the brain processes information one of two ways: either centrally (systematically), thoughtfully scrutinizing the arguments and evidence presented, or peripherally (heuristically) based on emotional, visual, and social cues. Whether a person processes content centrally or peripherally is contingent on his/her motivation and ability to do so. The social media environment itself (with huge quantities of bite-sized information, constantly updated in real time) promotes peripheral/heuristic processing by placing constraints on both users' ability and motivation to think carefully about the information presented. As a result, factors such as “who shared it” and “does this make me feel good/right” become more likely to dictate processing goals and dissemination behaviors.

Meanwhile, people with certain psychological traits or worldviews are especially likely to rely on heuristics in deciding what to think about - and do with - news stories. Traits such as conspiratorial ideation (Uscinski, Klofstad, Atkinson, 2016), as well as specific epistemic beliefs (faith in one's own intuition and belief that facts are politically constructed) (Garrett & Weeks, forthcoming), both of which transcend political ideology, are especially predictive of one's likelihood of embracing misinformation as true.

Yet, the impact of these traits is not fixed. The tendency to reflexively embrace conspiracy theories or judge stimuli based on “gut instinct” alone *can* be altered through thoughtful interventions and informational cues (Ecker et al., 2010;

Lewandowsky et al., 2012; Tully & Vraga, 2017; Uscinski et al, 2016). Tagging information as problematic, unverified, or stemming from an unreliable source increases users' motivation to centrally process the message rather than relying on heuristics and biases. As Garrett and Weeks (forthcoming) suggest, reminding audiences about the importance of evidence and the risks of being guided by feelings might be particularly effective when dealing with audiences who place faith in their own intuition and see truth as socially constructed.

### Discussion

There is no single intervention that can solve the problems created by the spread of false information in politics. But decades of research into how people attend to and evaluate information can shed light into how and why people process, believe, and share misinformation. By building upon this research to design effective evidence-based interventions, we believe it is possible to substantially reduce the impact of misinformation on political beliefs and attitudes.

### References

- Barnhurst, K. G., & Mutz, D. (1997). American journalism and the decline in event-centered reporting. *Journal of Communication*, 47, 27-53.
- Bennett, W. L. (2012). The personalization of politics: Political identity, social media, and changing patterns of participation. *The ANNALS of the American Academy of Political and Social Science*, 644(1), 20-39.
- Berinsky, A. J. (2015). Rumors and health care reform: Experiments in political misinformation. *British Journal of Political Science*, 47, 241-262.
- Brotherton R, Eser S. Bored to fears: Boredom proneness, paranoia, and conspiracy theories. *Personality and Individual Differences*. 2015;80:1-5. 19.
- Brotherton, R., French, C. C., & Pickering, A. D. (2013). Measuring belief in conspiracy theories: The generic conspiracist beliefs scale. *Frontiers in psychology*, 4.
- Cappella, J. N., & Jamieson, K. H. (1997). *Spiral of cynicism: The press and the public good*. New York: Oxford University Press.
- Dagnall N, Drinkwater K, Parker A, Denovan A, Parton M. Conspiracy theory and cognitive style: a worldview. *Frontiers in Psychology*. 2015;6(206). 20.
- Ecker, U. K. H., Lewandowsky, S., Chang, E. P., & Pillai, R. (2014). The effects of subtle misinformation in news headlines. *Journal of Experimental Psychology: Applied*, 20(4), 323-335.
- Ecker, U. K. H., Lewandowsky, S., & Tang, D. T. W. (2010). Explicit warnings reduce but do not eliminate the continued influence of misinformation. *Memory & Cognition*, 38, 1087-1100.
- Fischer, P., Schulz-Hardt, S., & Frey, D. (2008). Selective exposure and information quantity: How different information quantities moderate decision maker's preference for consistent and inconsistent information. *Journal of Personality & Social Psychology*, 94, 231-244.

- Garrett, R. K., Nisbet, E. C., & Lynch, E. K. (2013). Undermining the corrective effects of media-based political fact checking? The role of contextual cues and naïve theory. *Journal of Communication*, 63(4), 617-637.
- Garrett, R.K., & Weeks, B.E. (forthcoming). Epistemic beliefs' role in promoting misperceptions and conspiracist ideation. Proceedings of the National Academy of Sciences.
- Hopmann, D. N., Shehata, A., & Strömbäck, J. (2015). Contagious media effects: How media use and exposure to game-framed news influence media trust. *Mass Communication and Society*, 18, 776-798
- Jones, D. A. (2004). Why Americans don't trust the media. *The International Journal of Press/Politics*, 9, 60-75.
- Kantar Media (2016). *Brand and trust in a fragmented news environment*. Reported prepared for the Reuters Institute for the Study of Journalism.
- Klar, S., & Krupnikov, Y. (2016). *Independent politics*. Cambridge University Press.
- Lewandowsky, S., Gignac, G. E., & Oberauer, K. (2013). The role of conspiracist ideation and worldviews in predicting rejection of science. *PloS one*, 8(10), e75637.
- Lewandowsky S, Ecker UKH, Seifert CM, Schwarz N, Cook J. (2012) Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest*. 2012;13(3):106–131.
- Marder, B., Slade, E., Houghton, D., & Archer-Brown, C. (2016). I like them, but won't 'like' them: An examination of impression management associated with visible political party affiliation on Facebook. *Computers in Human Behavior*, 61, 280-287.
- Messing, S., & Westwood, S. J. (2014). Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. *Communication Research*, 41(8), 1042-1063.
- Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of communication*, 60(3), 413-439.
- Neuman, W. R. (2016). *The digital difference: Evolving media technology and the theory of communication effects*. Cambridge, MA: Harvard University Press.
- Nyhan, B., & Reifler, J. (2015). Displacing misinformation about events: An experimental test of causal corrections. *Journal of Experimental Political Science*, 2(1), 81–93.
- Shin, J., & Thorson, K. (2017). Partisan Selective Sharing: The Biased Diffusion of Fact-Checking Messages on Social Media. *Journal of Communication*.
- Tully, M., & Vraga, E. K. (2017). Effectiveness of a news media literacy advertisement in partisan versus nonpartisan online media

contexts. *Journal of Broadcasting & Electronic Media*, 61(1), 144-162.

Uscinski JE, Klofstad C, Atkinson MD.  
What Drives Conspiratorial Beliefs?  
The Role of Informational Cues and  
Predispositions. *Political Research  
Quarterly*. 2016;69(1):57–71.

Vraga, E. K., Thorson, K., Kligler-  
Vilenchik, N., & Gee, E. (2015).  
How individual sensitivities to  
disagreement shape youth political  
expression on Facebook. *Computers  
in Human Behavior*, 45, 281-289.

Watts, M. D., Domke, D., Shah, D. V., &  
Fan, D. P. (1999). Elite cues and  
media bias in presidential  
campaigns: Explaining public  
perceptions of a liberal press.  
*Communication Research*, 26, 144–  
175.

## Misinformation and Identity-Protective Cognition

Dan M. Kahan, Yale University

**Abstract:** This paper synthesizes existing work on misinformation relating to policy-relevant facts. It argues that misinformation has the greatest power to mislead when it interacts with identity-protective cognition.

### Introduction: IPC → misinformation

This paper furnishes a compact (even subcompact<sup>i</sup>) synthesis of the literature on miscommunication of decision-relevant science. The incidence and effect of such information, the paper argues, is conditional on identity-protective cognition on the part of its consumers (Nyhan & Reffler 2015; Flynn, Nyhan & Reifler 2017).<sup>ii</sup> The paper thus begins with a short discussion of identity-protective cognition before addressing the psychological dynamics of misinformation.

### What is IPC?

Identity-protective cognition (IPC) is a species of motivated reasoning. Motivated reasoning refers to the unconscious tendency of people to selectively credit and dismiss factual information in patterns that promote some goal or interest independent of the truth of the asserted facts. When that goal or interest is the protection of one's status within an important affinity group, the information consumer can be said to be displaying IPC (Sherman & Cohen 2006).

Empirical studies of the role IPC plays in public conflicts over science abound (Flynn et al. 2017; Kahan 2010). In one study (Kahan, Jenkins-Smith & Braman 2011), for example, subjects evaluated whether three scientists of impeccable credentials were knowledgeable and trustworthy “experts” on climate change, gun control, and nuclear waste disposal, respectively. The subjects’ evaluations, it turned out (e.g., Figure 1) depended strongly on the relationship between the scientists’ (experimentally manipulated) positions on those issues and the subjects’ own cultural commitments: if a scientist’s position matched the one that prevailed within the subjects’ cultural community, he was deemed an expert; if not, then not.

If this is a good model of how people identify scientific “experts” in the real world, then we should expect people of diverse cultural outlooks to be as polarized over what expert consensus is as they are over what the facts on politically charged issues are. This turns out to be true, the same study and countless others (e.g., Funk & Kennedy 2016) have shown.

Figure 1. Identity-protective reasoning in recognition of expertise. Adapted from Kahan, Jenkins-Smith & Braman (2011). Study subjects view highly credentialed scientist as an “expert” on climate change much less readily when the scientist takes the view opposed to rather than supportive of the one that is dominant in their political group. Colored bars reflect 0.95 level of confidence.

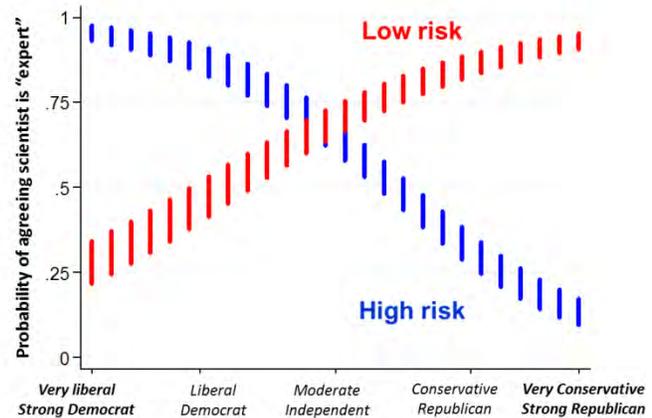
***Is this scientist an “expert” on global warming?***



Robert Linden  
 Position: Professor of Meteorology,  
 Massachusetts Institute of Technology  
 Education: Ph.D., Harvard University  
 Memberships:  
 • American Meteorological Society  
 • National Academy of Sciences

**“High risk”**  
 “It is now beyond reasonable scientific dispute that human activity is causing ‘global warming’ and other dangerous forms of climate change. . . .”

**“Low risk”**  
 “Judged by conventional scientific standards, it is premature to conclude that human CO<sub>2</sub> emissions—so-called ‘greenhouse gasses’—cause global warming. . . .”

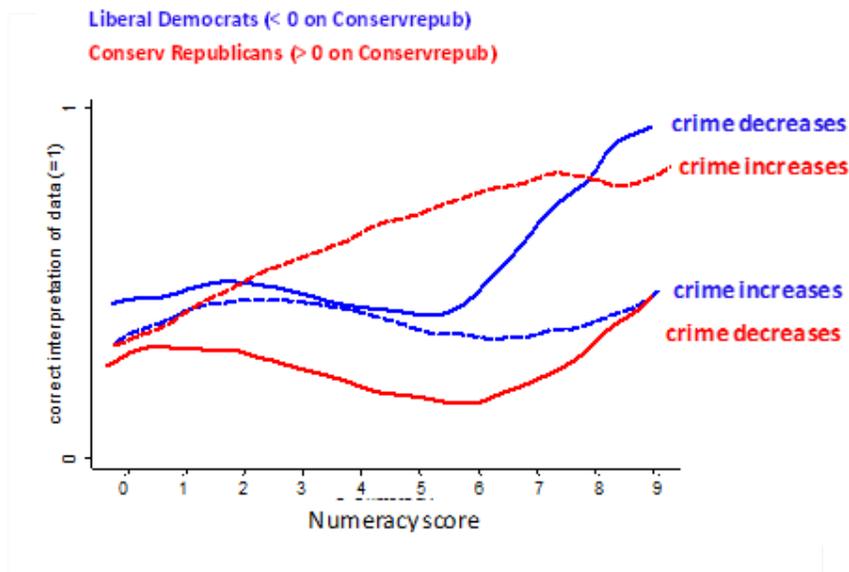


Surprisingly, IPC has been found to intensify, not abate, as the members of the public become more proficient in the forms of critical reasoning essential to science comprehension (Giner-Sorolla, & Chaiken 1997; Kahan 2016). *Numeracy* is an aptitude to reason well with quantitative information and to draw appropriate inferences from data (Peters et al. 2006). Thus, the individuals who are highest in numeracy are the best able to recognize whether evidence from a controlled experiment displays the pattern of covariance that supports or negates a hypothesis—*unless* that evidence relates to a

politically charged issue (e.g., gun control). In *that* case (Figure 2), the most numerate people are even more likely than the least numerate ones to construe such evidence as supporting the factual beliefs that prevail among people who share their political identity no matter what its true import (Kahan, Peters, Dawson & Slovic 2017).

If this is how people reason outside the lab, then we should expect people who are the most scientifically literate to be the most polarized on disputed forms of decision-relevant science. They are (Kahan 2015).

Figure 2. Responses by subjects of opposing cultural outlooks. Locally weighted regression lines (Cohen et al., 2003) track the proportion of subjects solving the problem correctly in relation to numeracy levels in the various experimental conditions. Blue lines plot relationships for subjects who score below the mean and red ones are for subjects who score above the mean on Conserv\_Repub, the composite measure based on liberal–conservative ideology and identification with one or the other major party. Solid lines are used for subjects in the condition in which the data, properly interpreted, support the inference that either skin rashes or crime decreased in response to an experimental treatment; dashed lines are used for subjects in conditions in which the data, properly interpreted, support the inference that either skin rashes or crime increased.



### IPC and dynamics of misinformation

Social scientists have documented a wide array of *misinformation* dynamics. Practically all of them bear the signature of IPC. Here are a few examples.

**Self-misinformation.** To begin, the consequences of misinformation must be assessed in relation to the disposition of culturally diverse citizens to *misinform themselves*. Consider the two experiments discussed so far. In the scientific-consensus experiment (Kahan et al 2011), no misinformation was necessary to induce subjects to reason in a manner that thwarted appropriate updating of their perceptions of what expert scientists believe. The same goes for the covariance-detection experiment: high-numeracy subjects were lulled into error by their own predispositions—not by the importuning of misinformers (Kahan et al. 2017). There is

thus good reason to believe that *even without* significant quantities of misinformation, we’d still see substantial polarization over decision-relevant science relating to culturally contested issues.

**Information search.** There is definitely a correlation between exposure to deceptive information in the media and being misinformed on a variety of consequential issues (e.g., Feldman et al. 2012). But in which direction does causation run? A considerable body of research concludes that people’s cultural and political predispositions are the source, not the outcome, of the information they consume. Identity-protection, not correctness, is their goal, here as well: armed with evidence, people are less vulnerable to succumb to opposing arguments (e.g., Arceneauz & Johnson 2013).

***Biased assimilation and polarization.***

*Biased assimilation* refers to the tendency of people to more readily credit information that reinforces their existing factual beliefs on controversial issues (e.g., the deterrent effect of capital punishment), thereby magnifying the gap between those on both sides of a dispute (Lord, Ross & Lepper 1979).

What IPC adds is an account of the *advent*, *direction*, and *incidence* of this information-processing distortion. Biased assimilation doesn't generate a clean prediction, for example, on whether and how individuals will react to information on a novel risk source (such as nanotechnology [Kahan, Braman, Slovic, Gastil & Cohen 2009] or the HPV vaccine [Kahan, Braman, Cohen, Gastil & Slovic 2010]). IPC, however, correctly predicts that even in this context members of opposing cultural groups will polarize (or not) based on tacit cues, such as in-group and out-group conflict (Kahan, Braman et al. 2010) or exposure to social memes that fuse the new risk source to existing ones on which there is already cultural polarization (Kahan, Jamieson, Landrum & Winneg 2017).

Biased assimilation also offers no explanation of why people will sometimes change their minds about policy-consequential facts. IPC does. The answer is not exposure to new, fact-laden, empirical arguments but rather personal observation of behavior by members of their own cultural group. People evince confidence in decision-relevant science when they rely on it to make important decisions. Such behavior vouches not only for the safety (or danger) of a putative risk source (novel or well-established) but also for the safety (or not) of holding the new position within their particular affinity group (Kahan 2015).

***Retrenchment and Backfiring.*** Attempts to correct factual misimpressions not only fail to move individuals. Perversely, the correction itself is viewed as certifying that the facts being challenged are within the constellation of beliefs that divide opposing groups. Responding to this cue, those who initially credited the misinformation for identity-protective purposes only dig in deeper when admonished that the beliefs in question are false (Nyhan, Reifler & Lubel 2013). Such “backfiring” has been observed to influence mass opinion on a diverse set of issues, from the safety of vaccines (Nyhan, Reifler, Richey & Freed 2014) to the existence of weapons of mass destruction in Iraq (Nyhan & Reifler 2010) to the consensus of expert views on climate change (Cook & Lewandowsky 2016).

**What about “fake news”?**

The general election of 2016 was inundated with instances of “fake news”—on-line fictional accounts disguised to resemble news stories published in legitimate newspapers and like media. Reasoning axiomatically from general dynamics of belief formation, commentators have tended to identify public consumption of “fake news” as an important part of Donald Trump's victory over Hilary Clinton in the electoral contest for President of the United States (e.g., Parkinson 2016).

But the only empirical study published so far casts considerable doubt on this conclusion. Combining an internet database of fake news stories with survey respondents' recollection of having seen particular articles, Allcott and Gentzkow (2017) report that individuals were much more likely to represent that they had read, agreed with, and shared articles that were favorable toward their preferred political candidate or unfavorable toward that candidate's opponent. This is evidence, then,

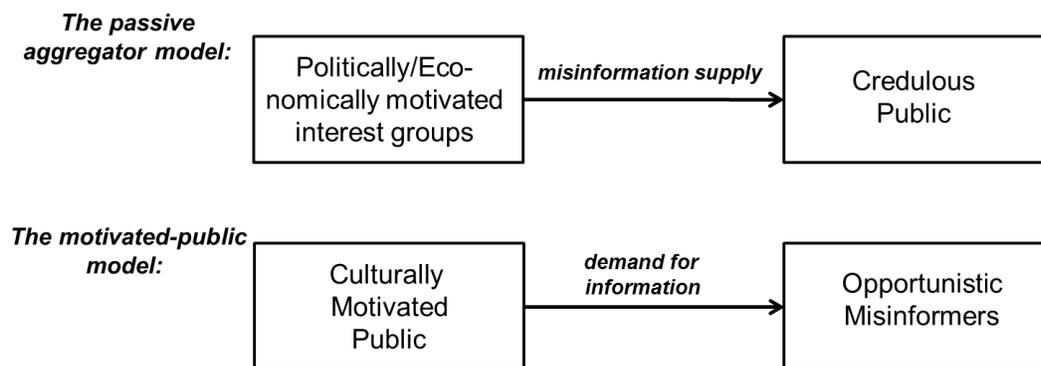
that consumption of the stories (or story: the average number of fake news articles read by a voter, A&G calculated, was 1) were a consequence, not a cause of their readers' voting preferences.

### Conclusion: Two models and one yuuuuge open question

Consider two models of the nature and impact of misinformation (Figure 3). The first account treats the public as credulous consumers of misinformation manufactured

by economic or political interests and directed at the public by a compliant or even complicit media. The second, in contrast, views the public as *motivated consumers* who consciously seek out information supportive of factual beliefs characteristic of their identity-defining cultural groups. That demand, according to the model, creates opportunities, economic and political, for conflict entrepreneurs to supply *misinformation* that the public will greedily consume as a result of IPC.

Figure 3. Two models of information processing.



The primary conclusion of this synthesis paper is that the latter, “motivated public” position is much closer to being true than is the former, “credulous public” one. Only the “motivated public” account explains the interaction between misinformation and the panoply of reasoning defects associated with identity-protective reasoning (IPC)—from self-deception to biased assimilation, from motivated system 2 reasoning to cultural backlash effects. It is this link to IPC that enables the “motivated public” account to explain why misinformation tends to amplify polarization rather than increase public misunderstandings generally.

The “motivated public” account also highlights the most important missing piece of existing research: how and why particular

positions on disputed facts acquire the antagonistic social meanings that turn them into badges of group identity (Lessig 1995; Kahan, Jamieson et al. 2017). Along with accident and misadventure, misinformation no doubt plays a role. But for this purpose, misinformation that takes aim only at the facts that underlie a disputed political issue will not be particularly effective; what will be is misinformation that implies the issue is one that already starkly divides members of opposing identity-defining groups; that’s the message that has the greatest potential to spark the sort of group-status conflict that disables the capacity of diverse citizens to recognize what science knows. Not until *that* sort of information is brought under control will it be possible to protect the

science communication environment from the particular toxin of IPC.

## References

- Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *J. Econ. Perspectives*, 31, 211-236.
- Arceneaux, K. & Johnson, M. (2013) *Changing Minds Or Changing Channels?: Partisan News in an Age of Choice* (University of Chicago Press).
- Cook, J. & Lewandowsky, S. (2016). Rational Irrationality: Modeling Climate Change Belief Polarization Using Bayesian Networks. *Topics in Cognitive Science* 8, 160-179.
- Feldman, L., Maibach, E.W., Roser-Renouf, C. & Leiserowitz, A. Climate on Cable. (2012) *The International Journal of Press/Politics* 17, 3-31.
- Flynn, D.J., Nyhan, B. & Reifler, J. (2017). The Nature and Origins of Misperceptions: Understanding False and Unsupported Beliefs About Politics. *Political Psychology* 38, 127-150.
- Funk, C. & Kennedy, B. (Oct. 4, 2016) *The Politics of Climate* (Pew Research Center).
- Giner-Sorolla, R. & Chaiken, S. (1997) Selective Use of Heuristic and Systematic Processing Under Defense Motivation. *Pers Soc Psychol B* 23, 84-97.
- Kahan, D. (2010). Fixing the Communications Failure. *Nature* 463, 296-297.
- Kahan, D. (2017) *Misconceptions, Misinformation, and the Logic of Identity-protective Cognition*. AAAS: Face of Science White Paper.
- Kahan, D., Braman, D., Cohen, G., Gastil, J., & Slovic, P. (2010). Who Fears the HPV Vaccine, Who Doesn't, and Why? An Experimental Study of the Mechanisms of Cultural Cognition. *Law and Human Behavior*, 34(6), 501-516.
- Kahan, D.M. Climate-Science Communication and the Measurement Problem. (2015). *Advances in Political Psychology* 36, 1-43.
- Kahan, D.M., Braman, D., Slovic, P., Gastil, J. & Cohen, G. (2009). Cultural Cognition of the Risks and Benefits of Nanotechnology. *Nature Nanotechnology* 4, 87-91.
- Kahan, D.M., Jamieson, K.H., Landrum, A. & Winneg, K. (2017). Culturally antagonistic memes and the Zika virus: an experimental test. *J Risk Res* 20, 1-40.
- Kahan, D.M., Jenkins-Smith, H. & Braman, D. (2011) Cultural Cognition of Scientific Consensus. *J. Risk Res.* 14, 147-174.
- Kahan, D.M., Peters, E., Dawson, E.C. & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy* 1, 54-86.

- Lessig, L. (1995). The Regulation of Social Meaning. *U. Chi. L. Rev.*, 62, 943-1045.
- Lord, C.G., Ross, L. & Lepper, M.R. (1979). Biased Assimilation and Attitude Polarization - Effects of Prior Theories on Subsequently Considered Evidence. *Journal of Personality and Social Psychology* 37, 2098-2109.
- Nyhan, B. & Reifler, J. (2015). The roles of information deficits and identity threat in the prevalence of misperceptions.
- Nyhan, B. & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Polit Behav* 32, 303-330.
- Nyhan, B., Porter, E., Reifler, J. & Wood, T. (2017). Taking Corrections Literally But Not Seriously? The Effects of Information on Factual Beliefs and Candidate Favorability. Available at <https://ssrn.com/abstract=2995128>.
- Nyhan, B., Reifler, J. & Ubel, P.A. (Nov. 27, 2016). The Hazards of Correcting Myths About Health Care Reform. *Medical Care* 51, 127-132  
110.1097/MLR.1090b1013e318279486b (2013).
- Parkinson, H.J. (2016) Click and elect: How fake news helped Donald Trump win a real election. *The Guardian*, available at <https://goo.gl/bL6Ww8>.
- Peters, E., Västfjäll, D., Slovic, P., Mertz, C.K., Mazzocco, K. & Dickert, S. Numeracy and Decision Making. (2006). *Psychol Sci* 17, 407-413.
- Sherman, D.K. & Cohen, G.L. (2006). The Psychology of Self-defense: Self-Affirmation Theory. *Advances in Experimental Social Psychology* 38, 183-242.

---

<sup>i</sup> For something approaching a mid-size account, see Kahan (2017).

<sup>ii</sup> If you want a luxury-sedan sized account of the sources and consequences of political misinformation, then get hold of these (e.g., Flynn, Nyhan & Reifler 2017; Nyhan, Porter et al 2017) and other papers by Nyhan and his collaborators.

## Using Behavioral Theory to Curb Misinformation Sharing

**Brian Southwell**, RTI International; Duke University; University of North Carolina  
**Vanessa Boudewyns**, RTI International

How can we stop the spread of misinformation? That question implies we have reason to be concerned about audience encounters with misinformation, a claim for which we have justification (Southwell & Thorson, 2015; Southwell, Thorson, & Sheble, 2018). Perhaps less obvious in that question but nonetheless crucial is exactly what spreading entails, including processes of human decision-making and behavior. Although automated social media accounts (or bots) discover and repost false material online (Maréchal, 2016) and algorithm-based approaches to countering misinformation offer some potential for remedy (e.g., Nguyen et al., 2013, or Bode & Vraga, 2015), the problem of misinformation spreading via electronic media outlets and through social networks is attributable to human behavior. Decades of academic concern with rumor suggest that the existence of falsehood alone is not solely responsible for false information diffusion, e.g., Allport and Postman (1947), Rosnow (1991), and Weeks and Southwell (2010); human psychology interacts with an information environment to disseminate falsehoods. If we consider conversation (and information exchange more broadly) between people to be fundamentally an observable *behavioral* phenomenon constrained by human motivations and skills and with important consequences (Southwell & Yzer, 2007), then misinformation sharing should be a function of human hopes and desires. By foregrounding misinformation sharing as behavior, we can both articulate important distinctions in the circumstances

in which sharing occurs and outline considerations for future intervention. A behavioral approach to understand misinformation sharing highlights at least three important variables with predictive and ethical consequence: awareness (of misinformation as such), ability (to share), and motivation (to act). Available literature suggests important limitations in human awareness of information accuracy, highlighting the potential for some unintentional spreading to occur and the potential utility of improving awareness. Considerations of how we can share information point to important roles for one's facility with information sharing technologies and circumstances (including emotionally-charged situations). Lastly, we can understand volitional sharing of misinformation through the lens of extant behavioral theory that has been widely applied to a range of civic and health-related behaviors, a move that highlights a path for future intervention involving more than simple awareness-raising.

### **Awareness and detection of misinformation**

If we plan to ask people to opt to not share inaccurate information, we first must assume their ability to recognize the existence of misinformation. Our task is inherently limited in the light of available cognitive psychology literature. As Southwell and Thorson (2015) argued, we can trace questions about our ability to detect misinformation back at least hundreds of years to a debate that unfolded across the roughly contemporaneous lives of

philosophers Descartes and Spinoza as they disagreed about what happens when people encounter false information (or any information). Rather than accepting Descartes' notion that people screen out false information at the point of exposure, Spinoza argued that people accept information by default and verify or reject it in a subsequent process. Empirical support for Spinoza's account has accumulated in recent decades (e.g., Asp et al., 2012; Gilbert, Krull, & Malone, 1990; Skurnik, Yoon, Park, & Schwarz, 2005), suggesting that people encode all new information as if it were true, even if only momentarily, and later tag the information as being either true or false. Despite this evidence of audience processing, Boudewyns and colleagues (2018) have noted that past research in domain-specific literatures on topics such as consumer advertising largely elides the question of individual awareness of false claims. The health advertising literature, for example, largely has focused on documenting problematic claims more than on audience awareness of false or misleading claims, *per se*. Such emphasis on the prevalence of misinformation to the exclusion of audience awareness research is unwarranted.

The pre-existing belief systems and social networks of audience members also constrain rejection of misinformation. Humans are vulnerable to deception in advertising, news stories, and even in interpersonal contexts in part because of our information processing tendencies, e.g., see Gilbert et al. (1993) for discussion. People tend to believe information that is compatible and consistent with their current beliefs, if the individual elements which comprise the information are coherent and without internal contradictions, if the source is perceived to be credible, and if there is a perceived social consensus (Lewandowsky

et al., 2012). Familiarity with a topic and a predisposition to scrutinize and distrust information may offer some protection, but assessing the credibility of information is not an easy task, especially if information is shared by people in your network (Benkler et al., 2017). People also can misattribute information sources when recalling past encounters with information, e.g., Auslander et al. (2017). Being connected largely to like-minded people can lead to echo chambers and filter bubbles that undermine identification of falsehoods (Pariser, 2011).

### **The various means of sharing**

Consider a false news story about a public official being an alcoholic planted by people opposing that official. Drawing on a recent catalogue of information-sharing behaviors (Southwell, 2013), there are various ways that any one person can share aspects of that story with others. A person could talk with friends, simply forward the story to another by e-mail, or post it on a bulletin board. A person might publicly protest or denigrate the story, either by declaring the presumed true story to be offensive to society or by claiming the person writing the story was lying; either way, the protest would share a form of the story with other people. A person also could coopt the original story to create a revised or new message; an evening talk show host making a joke or a person writing satire inspired by the story would further spread the story in a sense. Technology also now facilitates overt endorsement of a story; consider the explicit acknowledgement that a person approves a message by "liking" a post on Facebook or otherwise noting their endorsement on a web site by clicking on an icon.

Outlining these various forms of information sharing underscores the potential for unintentional sharing. Relatively inept users of social media technology, for example,

might accidentally share misinformation because of an errant click. On a different plane, a sarcastic comment to another person to mock a story known to be false nonetheless might raise awareness of the misinformation in a way not considered deeply by the speaker. New technologies also have facilitated information sharing and have introduced potential influences on misinformation sharing. With sharing now possible with a wide network almost immediately after reading content online, emotional arousal may now play a crucial role in that we have lowered the threshold of skills and ability necessary for sharing: misinformation spread is possible with an anxious tap on a keyboard.

### **Misinformation sharing as human behavior**

Following Southwell and Yzer (2007)'s conceptualization of conversation and Cappella's (1987) definition of interpersonal interaction, we contend that misinformation sharing is an observable *behavioral* phenomenon involving multiple people constrained by human motivations and skills and with important consequences. That framing sets the stage for future research involving beliefs and perceived incentives about information sharing that could extend beyond studies on public understanding of particular facts.

What are the beliefs and perceived incentives that appear to drive information and misinformation sharing? Available literature on news sharing and word-of-mouth diffusion offers useful evidence (Lee & Ma, 2012; Berger, 2014; Chen et al., 2015). Berger (2014), for example, suggests that word of mouth serves five key functions: impression management, emotion regulation, information acquisition, social bonding, and persuasion. These mirror beliefs that people can hold about sharing

misinformation. For example, people may think that sharing misinformation will generate social support or help them take vengeance (an aspect of emotion regulation). People may think sharing misinformation will help to communicate a specific identity (which relates to impression management) or reinforce shared views (social bonding). Sharing misinformation might even help a person gain new information from others. Each of these factors could encourage misinformation sharing even in situations in which a person understands the particular information they are sharing is false.

A range of theoretical frameworks that explain and predict social behavior are useful here. The Reasoned Action approach to predicting and explaining behavior, for example, holds important implications for our specific quest to understand underlying drivers behind a person's intentions to share potentially false information (Fishbein & Ajzen, 2010; Fishbein & Cappella, 2006; Fishbein & Yzer, 2003). A principal tenet of that approach – that intentions most proximally influence behavior – suggests that for behavior to change, interventionists must target precursors of intention, including attitudes concerning the behavior (e.g., approval, disapproval, neutrality, disinterest), beliefs about subjective norms concerning the behavior (how do others think one should respond?), and perceptions about one's confidence in performing the behavior. The more a person believes that sharing misinformation will result in positive consequences, that people who are important to him or her think he or she should share the misinformation (and that the norm is to share such information), and that they are confident in their ability to share the misinformation, the more likely they are to do so. Conversely, conscious effort to avoid sharing is likely a function of a similar array of perceptions.

Environmental and intrapersonal variables beyond individual beliefs also are noteworthy. Misinformation must be available in some form to be shared; the environmental prevalence of misinformation is a precondition for its spread (unless a person simply generates misinformation on their own). In addition, we must also consider individual variation (between people and across time within a person) in emotion and physiology. Misinformation sharing may help a person to regulate their emotions (Berger, 2014) and might also prompt certain emotions, like pride or guilt, that could offer incentive or disincentive for sharing. Emotion can be accommodated within existing cognitive frameworks such as the Reasoned Action approach. For example, if a person believes that sharing misinformation will allow them to vent (which they perceive as positive), then that belief will contribute to a positive attitude. Competing emotions might also discourage sharing, e.g., sharing misinformation may make someone feel guilty.

### **Implications for intervention and future research**

Expecting people to detect false information on their own and to avoid sharing it seems unwise; evidence suggests people often are vulnerable to accepting misinformation at face value. Similarly, simply asking people to not spread misinformation is unlikely to stop sharing in all cases. Communication technology makes some instances of inadvertent sharing possible, and such accidents could be reduced by adding a confirmation step before sharing occurs. Volitional sharing of misinformation may pose a greater challenge. Addressing such sharing will require that we understand the social identity, bonding, and emotion regulation benefits of sharing. Future survey research could elicit beliefs about misinformation sharing and researchers

could use those beliefs to build predictive models. With that work in hand, we may be in a better position to respectfully offer alternative ways for people to satisfy key needs without spreading misinformation.

### **References**

- Allport, G.W., & Postman, L. (1947). *The psychology of rumor*. New York: Holt, Rinehart & Winston.
- Asp, E., Manzel, K., Koestner, B., Cole, C. A., Denburg, N. L., & Tranel, D. (2012). A neuropsychological test of belief and doubt: Damage to ventromedial prefrontal cortex increases credulity for misleading advertising. *Frontiers in Neuroscience*, 6, 1–9.
- Auslander, M. V., Thomas, A. K., & Gutches, A. H. (2017). Confidence moderates the role of control beliefs in the context of age-related changes in misinformation susceptibility. *Experimental Aging Research*, 43(3), 305-322.
- Benkler, Y., Faris, R., Roberts, H., & Zuckerman, E. (2017 March 3). Study: Breitbart-led right-wing media ecosystem altered broader media agenda. *Columbia Journalism Review*. Retrieved from <https://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php>.
- Berger, J. (2014). Word of mouth and interpersonal communication: A review and directions for future research. *Journal of Consumer Psychology*, 24(4), 586-607.
- Bode, L., & Vraga, E. K. (2015). In related news, that was wrong: The correction of misinformation through

- related stories functionality in social media. *Journal of Communication*, 65(4), 619–638.
- Boudewyns, V., Southwell, B., Betts, K., Gupta, C., Paquin, R., O'Donoghue, A., & Vazquez, N. (2018). Awareness of misinformation in health-related advertising: a narrative review of the literature. In B. Southwell, E. Thorson, & L. Sheble (eds.). *Misinformation and mass audiences*. Austin, TX: University of Texas Press. p. 35-50.
- Cappella, J. N. (1987). Interpersonal communication: Definitions and fundamental questions. In: C. R. Berger & S. H. Chaffee (eds.). *Handbook of communication science*. Newbury Park, CA: Sage. p. 184-238.
- Chen, X., Sin, S. C. J., Theng, Y. L., & Lee, C. S. (2015). Why students share misinformation on social media: Motivation, gender, and study-level differences. *The Journal of Academic Librarianship*, 41(5), 583-592.
- Fishbein, M., & Ajzen, I. (2010). *Predicting and changing behavior: the reasoned action approach*. New York: Psychology Press.
- Fishbein, M., & Cappella, J. N. (2006). The role of theory in developing effective health communications. *Journal of Communication*, 56(s1), s1-s17.
- Fishbein, M., & Yzer, M. C. (2003). Using theory to design effective health behavior interventions. *Communication Theory*, 13, 164-183.
- Gilbert, D. T., Krull, D. S., & Malone, P. S. (1990). Unbelieving the unbelievable: some problems in the rejection of false information. *Journal of Personality and Social Psychology*, 59, 601–613.
- Gilbert, D., Tafarodi, R., & Malone, P. (1993). You can't not believe everything you read. *Journal of Personality and Social Psychology*, 65(2), 221-233.
- Lee, C. S., & Ma, L. (2012). News sharing in social media: The effect of gratifications and prior experience. *Computers in Human Behavior*, 28(2), 331–339.
- Lewandowsky S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3), 106–131.
- Maréchal, N. (2016). When bots tweet: toward a normative framework for bots on social networking sites. *International Journal of Communication*, 10, 10. Retrieved from <http://ijoc.org/index.php/ijoc/article/view/6180/1811>.
- Nguyen, N. P., Yan, G., & Thai, M. T. (2013). Analysis of misinformation spread containment in online social networks. *Computer Networks*, 57(10):2133–2146.
- Pariser, E. (2011). *The filter bubble: what the internet is hiding from you*. London: Penguin UK.
- Rosnow, R.L. (1991). Inside rumor — A personal journey. *American Psychologist*, 46(5), 484–496.

- Skurnik, I., Yoon, C., Park, D. C., & Schwarz, N. (2005). How warnings about false claims become recommendations. *Journal of Consumer Research*, 31(4), 713–724.
- Southwell, B. G. (2013). *Social networks and popular understanding of science and health: sharing disparities*. Baltimore, MD: Johns Hopkins University Press.
- Southwell, B. G., & Thorson, E. A. (2015). The prevalence, consequence, and remedy of misinformation in mass media systems. *Journal of Communication*, 65(4), 589-595.
- Southwell, B. G., Thorson, E. A., & Sheble, L. (Eds.). (2018). *Misinformation and mass audiences*. Austin, TX: University of Texas Press.
- Southwell, B. G., & Yzer, M. C. (2007). The roles of interpersonal communication in mass media campaigns. In: C. Beck. (ed.). In *Communication yearbook 31*. New York: Lawrence Erlbaum Associates. p. 420-462.
- Weeks, B., & Southwell, B. (2010). The symbiosis of news coverage and aggregate online search behavior: Obama, rumors, and presidential politics. *Mass Communication and Society*, 13(4), 341-360.

## Assessing Current Efforts by the Platforms and their Effectiveness

Claire Wardle, First Draft, Shorenstein Center on Media, Politics and Public Policy, Harvard Kennedy School of Government

Over the past twelve months, ideas for solving mis- and disinformation have been discussed endlessly at conferences and in workshops, but we've seen few concrete changes and nothing has significantly moved the needle. While there is certainly more money available from the philanthropic sector, and a myriad of small projects are underway, grand ideas are yet to be implemented. While there are projects developing in terms of news literacy projects, third party technology initiatives, credibility scores, and indicators, this paper will focus on the steps taken by the social platforms.

### What have the social networks done to tackle agents motivated by financial gain?

As we have seen, one of the primary motivations for creating disinformation is financial gain. Both Google and Facebook have taken steps to prevent revenue flowing to the owners of "bad sites, scams and ads." Google permanently banned nearly 200 publishers from its AdSense advertising network as of late 2016 (Spencer 2017). Facebook made similar moves, updating its policies with language stating they would not display ads that show misleading or illegal content. Facebook has also taken steps to tackle 'imposter content,' stating, "On the buying side, we've taken action against the ability to spoof domains, which will reduce the prevalence of sites that pretend to be real publications" (Written Evidence). And, in late August 2017, Facebook announced they would block ads from pages that had repeatedly shared false news, stating "Currently, we do not allow

advertisers to run ads that link to stories that have been marked false by third-party fact-checking organizations. Now we are taking an additional step. If Pages repeatedly share stories marked as false, these repeat offenders will no longer be allowed to advertise on Facebook" (Shukla 2017).

Facebook moved relatively quickly from Mark Zuckerberg's post-election denial (Shahari 2017) that 'fake news' was a problem on his platform, to rolling out a third-party fact-checking initiative on December 15, 2016 that includes the International Fact Checking Network, *The Associated Press*, *The Washington Post*, and *Snopes* (Mosseri 2016). They expanded the project to France and Germany in February and the Netherlands in March. In this initiative, users flag posts they think might be 'false news,' populating a queue that the affiliated fact-checking organizations can see. In August, Facebook announced it would use "updated machine learning" to detect more potential hoaxes. After an article has been fact-checked, any user who sees that content will see that it has been disputed by one of the fact-checking organizations. If someone tries to share a disputed article, they are reminded with a pop-up notice that the content is in dispute. In December 2017, Facebook announced (Lyons 2017) it would replace these disputed flags with 'Related Articles,' a featured box underneath a post that was designed to "help people discover new articles they may find interesting about the same topic" (Su 2017).

After months of requests by the fact-checking organizations themselves (as well as researchers) for data connected to the initiative, in early October, Facebook provided information via an email (that was then reported on by Craig Silverman of BuzzFeed) that when content was flagged by third party fact-checkers sharing decreased by 80% (Silverman 2017).

### **What have the social networks done to tackle agents motivated by political influence?**

In April 2017, Facebook's Security team published a paper on 'Information Operations,' defining it as "actions taken by organized actors (governments or non-state actors) to distort domestic or foreign political sentiment, most frequently to achieve a strategic and/or geopolitical outcome" (Weedon, et al 2017). It was the first-time Facebook had admitted the scale of the problem they face in terms of official, tightly organized, networked agents using their platform to disseminate automated dis-information.

In September 2017, Facebook admitted that they had found evidence that 'dark ads' (ads that are only visible to the intended audience, rather than publicly viewable on a page) had been purchased by a Russian organization and directed at US citizens. Facebook explained, "[T]he ads and accounts appeared to focus on amplifying divisive social and political messages across the ideological spectrum — touching on topics from LGBT matters to race issues to immigration to gun rights" (Stamos 2017). A few days later, an investigation by the *Daily Beast* found inauthentic accounts, seemingly located in Russia, had used the Facebook events function to organize anti-immigration protests in the US (Collins, et al 2017).

Twitter has also come under scrutiny recently. In a June 2017 blog post, Twitter explained its efforts to fight against bots: "We're working hard to detect spammy behaviors at source,... [and we] also frequently take action against applications that abuse the public API to automate activity on Twitter (Twitter 2016), stopping potentially manipulative bots at the source" (Crowell 2017). They were forced to update this post in late September (Twitter 2017), summarizing the role the platform may have played in terms of Russian interference and the 2016 election, explaining some of the tools they are using to prevent automation, coordinated attacks, and campaigns on the platform. They also argued that researchers using Twitter's own data would not see the steps the platform was taking against this type of coordinated activity, as some of the enforcement would not be visible via the public API. While this is certainly true, the fact that researchers cannot independently assess the effectiveness of these moves makes it very difficult to understand this phenomenon (Warzel 2017).

One significant issue that has surfaced recently is the challenge for researchers who are attempting to understand this phenomenon when the platforms are deleting content. As Jonathan Albright explained (Timberg and Dwoskin 2017) to the Washington Post, his analysis of ads, pages, and posts connected to the 2016 elections was taken down by Facebook after he published his research, meaning that his methodology couldn't be replicated, and there was no opportunity for other researchers to examine patterns in the data to more fully understand the methods employed.

### **Providing additional context**

Google News recently took steps to allow publishers to highlight fact-checked content

for programmatic detection using schema.org, a structured data markup schema supported by the major search engines. This feature first appeared in the UK and the US last October, and in the summer of 2017 added the same functionality to Google News in Germany, France, Brazil, Mexico, and Argentina.

In early October 2017, Facebook started to test new contextual features. As they explained in their blog post announcing the move:

“For links to articles shared in News Feed, we are testing a button that people can tap to easily access additional information without needing to go elsewhere. The additional contextual information is pulled from across Facebook and other sources, such as information from the publisher’s Wikipedia entry, a button to follow their Page, trending articles or related articles about the topic, and information about how the article is being shared by people on Facebook. In some cases, if that information is unavailable, we will let people know, which can also be helpful context” (Anker, et al 2017).

This move is supported by experimental research by Leticia Bode in 2015 which suggested that when a Facebook post that includes mis-information is immediately contextualized in their ‘related stories’ feature underneath, misperceptions are significantly reduced (Bode and Vraga 2015). Hopefully, Facebook will be forthcoming with data about the success of the experiment, to provide cues for the other platforms about whether people will click on additional contextual information if provided with the option.

## Conclusion

As this article goes to print, Facebook has announced a significant change to their Newsfeed algorithm, announcing that posts from friends and family will appear higher than posts from news publishers and brands (Mosseri 2018). There was also a suggestion that journalism outlets will be ranked by independent assessments of credibility (Seetharaman 2018), for example via surveys of users in terms of which outlets they would be more prepared to pay for. The relative weighting of these elements within the algorithm is unclear. (For example, an outlet might have a low credibility score, but if all of your family and friends comment on a story from that outlet, will you still see it?)

When assessing moves by the technology companies over the past year, one of the most frustrating elements has been the failure to connect with the research, education, library, civil society, and policy communities at any substantive level. There are decades of research on mis-information, how people ‘read’ and make sense of information, and the factors that slow down or exacerbate rumors. But the responses have often felt knee-jerk and atheoretical, and at times public relations moves, rather than serious attempts to tackle the complexity of the problem. On this topic, when the scale and seriousness require sophisticated responses, the technology companies must work more closely with those who have research expertise on this subject, as well as those working on the ground around the world, and experience first-hand the real-world repercussions of polluted information streams.

## References

- Anker, A., S. Su, and J. Smith (October 5, 2017) News Feed FYI: New Test to Provide Context About Articles, Facebook Newsroom, <https://newsroom.fb.com/news/2017/10/news-feed-fyi-new-test-to-provide-context-about-articles/>
- Bode, L. & Vraga, E (2015) In Related News, That Was Wrong: The Correction of Mis-information Through Related Stories Functionality in Social Media, *Journal of Communication*, 65 (4): 619-638.
- Collins, B. et al (September 11, 2017) Exclusive: Russia Used Facebook Events to Organize Anti-Immigrant Rallies on U.S. Soil, *The Daily Beast*, <http://www.thedailybeast.com/exclusive-russia-used-facebook-events-to-organize-anti-immigrant-rallies-on-us-soil>
- Crowell, C., (June 14, 2017) Our Approach to Bots and Automation, [https://blog.twitter.com/official/en\\_us/topics/company/2017/Our-Approach-Bots-Mis-information.html](https://blog.twitter.com/official/en_us/topics/company/2017/Our-Approach-Bots-Mis-information.html)
- Lyons, T. (2017) News Feed FYI: Replacing Disputed Flags with Related Articles, Facebook Newsroom, <https://newsroom.fb.com/news/2017/12/news-feed-fyi-updates-in-our-fight-against-misinformation/>
- Mosseri, A. (2016) News Feed FYI: Addressing Hoaxes and Fake News, Facebook Newsroom, December 15, 2016.
- Mosseri, A. (2018) News Feed FYI: Bringing People Closer Together, Facebook Newsroom, <https://newsroom.fb.com/news/2018/01/news-feed-fyi-bringing-people-closer-together/>
- Seetharaman, D., Alpert, L., & Mullin, B. (2018) Facebook to Overhaul How It Presents News in Feed, *Wall Street Journal*, <https://www.wsj.com/articles/facebook-considers-prioritizing-trustworthy-news-sources-in-feed-1515714244>
- Shahari, A. (Nov 11. 2017) Zuckerberg Denies Facebook's Impact on the Election, <http://www.npr.org/sections/alltechconsidered/2016/11/11/501743684/zuckerberg-denies-fake-news-on-facebook-had-impact-on-the-election>
- Shukla, S. (August 28, 2017) Blocking Ads from Pages that Repeatedly Share False News. Facebook Newsroom. <https://newsroom.fb.com/news/2017/08/blocking-ads-from-pages-that-repeatedly-share-false-news/>
- Silverman, C. (2017) Facebook Says Its Fake News Label Helps Reduce The Spread Of A Fake Story By 80%, *Buzzfeed*, [https://www.buzzfeed.com/craigsilverman/facebook-just-shared-the-first-data-about-how-effective-its?utm\\_term=.ujVvrWlzn#.ba7mpJMP](https://www.buzzfeed.com/craigsilverman/facebook-just-shared-the-first-data-about-how-effective-its?utm_term=.ujVvrWlzn#.ba7mpJMP)
- Spencer, S. (Jan. 25, 2017) How we fought bad ads, sites and scammers in 2016, *Google Blog*, <https://www.blog.google/topics/ads/how-we-fought-bad-ads-sites-and->

scammers-2016

Stamos, A. (September 6, 2017) An Update On Information Operations On Facebook, Facebook Newsroom, <https://newsroom.fb.com/news/2017/09/information-operations-update/>

Su, S. (April 25, 2017) News Feed FYI: New Test With Related Articles, <https://newsroom.fb.com/news/2017/04/news-feed-fyi-new-test-with-related-articles/>

Timberg, C. & E. Dvoskin (Oct 12, 2017) Facebook takes down data and thousands of posts, obscuring reach of Russian disinformation, The Washington Post, <https://www.washingtonpost.com/news/the-switch/wp/2017/10/12/facebook-takes-down-data-and-thousands-of-posts-obscuring-reach-of-russian-disinformation/>

Twitter, (April 6, 2016) Automation Rules, <https://support.twitter.com/articles/76915>

Twitter Public Policy (Sept 28, 2017) Update: Russian Interference in 2016 US Election, Bots, & Misinformation, Twitter Blog, [https://blog.twitter.com/official/en\\_us/topics/company/2017/Update-Russian-Interference-in-2016--Election-Bots-and-Misinformation.html](https://blog.twitter.com/official/en_us/topics/company/2017/Update-Russian-Interference-in-2016--Election-Bots-and-Misinformation.html)

Warzel, C. (October 3, 2017) Researchers Are Upset That Twitter Is Dismissing Their Work On Election Interference, BuzzFeed, <https://www.buzzfeed.com/charliwarzel/researchers-are-upset-that->

twitter-is-dismissing-their-work?

Weedon, J., W. Nuland & A. Stamos (April 27, 2017) Information Operations and Facebook, p. 4 <https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf>

Written evidence by Facebook to the UK Parliamentary Inquiry on Fake News, <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/culture-media-and-sport-committee/fake-news/written/49394.html>

## **Information Operations and Democracy: What Role for Global Civil Society and Independent Media?**

**Sarah Oh**, The Center for Information Technology Research in the Interest of Society (CITRIS)  
and the Banatao Institute

There is increasing concern that the manipulation of social media<sup>i</sup> through targeted information operations (Weedon 2017) could play a significant role in influencing the outcome of future elections and other democratic events and processes globally (Deb 2017). Opinions about what can or should be done to address the production and spread of misinformation, disinformation, and false information, however, are divided (Anderson 2017). Stakeholders in the United States and Western Europe have begun testing a range of responses including regulatory remedies (Overview 2017), technology solutions to filter online content, and public education initiatives to promote information literacy around political events in their countries (Fleming 2016). These efforts, which are described as content risk support in this paper, have stemmed from discussions around the U.S., French, and German elections and rely on engagement with technology company representatives, foundations, think tanks, and researchers based in the U.S. or Western Europe.

Despite the focus on the experience in the West, concerns about information operations and their effect on democracy are salient in countries transitioning to democracy, where government and media institutions are under immediate threat or unevenly established, if at all. This paper draws on field experience<sup>ii</sup> in Eastern Europe and Southeast Asia to make the argument that more attention should be paid to groups engaged on this issue outside the West.

In Eastern Europe and Southeast Asia, there are three relevant trends around the aggressive use of social media to promote false information by the following groups: ISIS-aligned terror groups, domestic extremist nationalist and religious groups, and authoritarian governments, including Russia. Social media platforms present a new front for the ongoing ideological warfare that the Kremlin has waged in the Balkans and Baltic region (Priest 2017). In Southeast Asia, authoritarian governments themselves use social media to promote false information. In both regions, ISIS-aligned terror groups have strengthened their stronghold where there is evidence of increasing strength (Militant) and the use of social media to recruit supporters (Mejdini 2017).<sup>iii</sup> In both regions, extremist nationalist groups are flourishing online (Ristic 2017).

The effects of these trends may be greater than in the West for several reasons. In Indonesia, the Philippines, and Myanmar, the use of false information to influence public opinion is pernicious because of the widespread use of social media to access news and information; Facebook is a critical (if not the only) information lifeline. In Southeast Asia alone, there are more than 241 million Facebook users, which is approximately 12% of users worldwide (Connecting 2016); in Indonesia, Myanmar, and the Philippines, “Facebook is the internet.” Second, recent violence and conflict, and, significantly, the lack of institutions such as a functioning legal

system, independent media, and civil society mean there are less counter-weights in society to mitigate false and dangerous information campaigns.

For the reasons stated above, the perspective and data possessed by organizations in Southeast Asia and the Western Balkans is valuable. In the Philippines, Myanmar, and Indonesia, as well as Serbia, Kosovo, Albania, and Macedonia, civil society and independent media have been proactively studying and responding to actions taken by nationalist and extremist groups, and governments online around elections and other major political events. Increasingly, stakeholders familiar with these environments take the normative position that these groups should be engaged in discussions about possible defensive and offensive strategies against information operations particularly with industry leaders and academic researchers who seek to learn more about the manipulation or misuse of their platforms and advertising features<sup>iv</sup> globally.

Despite opportunities, however, discussions around operationalizing engagement between tech industry representatives and academic researchers with civil society and media organizations in transition countries are absent. Existing forums are inadequate for constructive engagement,<sup>v</sup> say local organizations, and there is a sense of fatigue and frustration among these organizations that may prevent their knowledge and insights from informing essential discussions in the future. There is a need to invest resources to develop these relationships explicitly during critical democratic transition periods.

### **An Emerging Community of Content Risk Researchers**

There are, at a minimum, three important functions civil society, media organizations, and other international groups could perform concerning content risk research and response. First, **civil society organizations** focused on issues such as citizen election observation, democracy, and citizen engagement are the front line of defense: users may call on or consult these organizations with social media problems, or with concerns about political events.<sup>vi</sup> **New independent media organizations** have studied information operations through reporting and analysis, which have ranged from anecdotal to in-depth reports and data analysis.<sup>vii</sup> Finally, **international practitioners**, who understand the context in which local organizations work, bring conflict sensitivity awareness and political development expertise useful for external actors. They also act as important facilitators when internal or community conflicts arise among organizations.

Collectively, the groups listed above possess unique knowledge about content risk trends in developing democracies. They often collect data about developments in real-time, and conduct analysis and share insights with local communities. They have identified trends such as the influence of ambient fear on why someone might share low-grade information, and have developed impactful, often creative, counter-messaging strategies. They also explain local cultural and political context around why a social media post or comment may or may not violate terms of service agreements. These insights are critical during moments of crisis.

### **Challenges for Engagement**

The unique knowledge civil society and independent media possess has not yet been harnessed; global civil society and media

has not yet fully occupied a position to advise industry and policy leaders on these issues on the global stage.<sup>viii</sup> In fact, existing engagement with civil society and media has fallen short because interventions have failed to address the challenges listed below.

First, groups may fear being viewed as conduits of international actors or companies. In some countries, say local groups, engagement with Facebook is seen negatively because of the company's coordination with government representatives. In addition, disagreements about Facebook content decisions have, at times, been wrongfully directed toward organizations by community members.<sup>ix</sup> In general, organizations working on issues related to democracy in developing democracies often work under significant pressure either from governments and security agents in their countries, other non-state actors, and even other civil society organizations. They may be subject to harassment themselves and often risk their livelihoods to do the work that they do.

Second, civil society and independent media organizations in developing democracies are young organizations that are still growing; they were formed after transitions or democracy movements that have occurred in recent history. Given the dearth of government funding and private foundations in transition countries, these young organizations rely on funding from international organizations and foundations and foreign governments, which means their funding is project-based. As a result, their capacity to conduct non-project related work is significantly constrained when compared to more established groups. Unless there is an explicit focus of their program to liaise with global policymakers and companies, these interactions cannot be prioritized or

proactively developed over providing direct project services, for example.

Third, major research institutions simply do not have direct relationships with international civil society and emerging media organizations covering the issues largely because of their own resource constraints, and lack of previous engagement in these countries.

Finally, language presents challenges. Representatives seeking to engage organizations and communities in transition countries primarily work in English, even if their platform serves localized content. Small, rights groups with valuable insights and knowledge who do not speak English face barriers to participate in discussions about policies that govern content on social media platforms.

### **Opportunities for Investigation**

Interviews conducted with civil society organizations in Southeast Asia and the Western Balkans suggest there are opportunities where research could shed light in three important areas.

*How should content risk be quantified in developing democracies?* How might context be better factored into risk assessment and identification when considering content trends and behaviors in transition countries (Adams et al)?<sup>x</sup> Are factors such as the state of media capture in a country, trust or distrust in government and media institutions, the role of social media in providing news, availability of independent media, digital and media literacy, active war or conflict, historical divisions in society, and other socio-economic factors that contribute to audience worldviews relevant? *Implications:* develop foundational risk identification categories that can be used by civil society, media, international researchers, and

industry partners to share an understanding of risk and investigate possible content issues.

***What organizational structures are needed to facilitate qualitative research on content risk?*** How might existing forums be improved to foster constructive engagement between technology companies and civil society and independent media organizations around future events, or post-mortem around past events? Between researchers and civil society and independent media? Can they be modified and used to discuss information integrity and threats to them? *Implications:* advise companies and thought leaders on structures for mutually beneficial engagement between local organizations in transition countries and companies, in addition to international researchers.

***What skills and resources do civil society organizations and global stakeholders need to engage more meaningfully with each other when it comes to content risk identification and management?*** What people or individuals might be useful for civil society organizations and technology companies to proactively identify and mitigate future content risk events? What interlocutors would enhance engagement between these two groups? Many local organizations note the importance of connectors, or individuals familiar with their environment and corporate representatives. How can they be surfaced and their capabilities enhanced? What capabilities are missing among technology companies and with civil society organizations? How might insights gained by those at social media companies interacting with civil society actors be mainstreamed across teams and organizations? *Implications:* inform the restructuring, or creation of new teams and hiring of new connectors who will seek to understand, assess, and mitigate risk in these

markets, and investigate possible terms of service violations.

To date, transition countries have not been given enough attention by researchers and tech industry representatives seeking to address the online information operations problem. Civil society and media organizations are on the front lines of responding to manifestations of this problem in Southeast Asia, the Western Balkans, and beyond. As social media companies are scrutinized, and the role of government is discussed, there is a larger question to raise around how civil society and media organizations can inform ongoing discussions. Without explicit efforts to engage civil society and media organizations in developing democracies, and use their knowledge, the tech industry and global policymakers are missing an opportunity. The dynamic nature of the problem makes identifying specific solutions challenging; developing relationships with a diverse set of invested stakeholders, or pathways for information to be shared to identify trends, and factor this knowledge into response strategies may be an important next step.

## References

- Adams, N. et al. “The Credibility Indicators Working Group.” Meedan.  
<https://meedan.com/credibility-indicators/>
- Anderson, J. and L. Rainie. “The Future of Truth and Misinformation Online.” Pew Research Center, October 19, 2017.  
<http://www.pewinternet.org/2017/10/19/the-future-of-truth-and-misinformation-online/>
- “Connecting Over 241 Million People in Southeast Asia on Facebook.” Facebook Business, March 15, 2016.

- <https://www.facebook.com/business/news/Connecting-over-241-Million-People-in-Southeast-Asia-on-Facebook>
- Deb, A., S. Donohue, and T. Glaisyer. "Is Social Media a Threat to Democracy?" The Omidyar Group, October 1, 2017. <https://www.omidyargroup.com/wp-content/uploads/2017/10/Social-Media-and-Democracy-October-5-2017.pdf>
- Fleming, J. "News Literacy and a Civics Model for Journalism Education." Center for News Literacy, 2016. [http://www.centerfornewsliteracy.org/2016\\_site/wp-content/uploads/2017/05/Jennifer-Fleming-FINAL.pdf](http://www.centerfornewsliteracy.org/2016_site/wp-content/uploads/2017/05/Jennifer-Fleming-FINAL.pdf)
- Mejdini, F. et al. "Balkan Jihadi Warriors Remain Safe on the Net." Resonant Voices Initiative, February 2, 2017. <http://www.balkaninsight.com/en/article/balkan-jihadi-warriors-remain-safe-on-the-net-01-27-2017>
- "Militant Islamists Shift Focus to Southeast Asia." *Financial Times*. <https://www.ft.com/content/b830552a-5176-11e7-bfb8-997009366969?mhq5j=e5>
- "Overview of the NetzDG Network Enforcement Law." CDT, July 17, 2017. <https://cdt.org/insight/overview-of-the-netzdg-network-enforcement-law/>
- Priest, D. and M. Birnbaum. "Europe Has Been Working to Expose Russian Meddling for Years." *Washington Post*, June 25, 2017. [https://www.washingtonpost.com/world/europe/europe-has-been-working-to-expose-russian-meddling-for-years/2017/06/25/e42dcece-4a09-11e7-9669-250d0b15f83b\\_story.html](https://www.washingtonpost.com/world/europe/europe-has-been-working-to-expose-russian-meddling-for-years/2017/06/25/e42dcece-4a09-11e7-9669-250d0b15f83b_story.html)
- Ristic, M. et al. "Far-Right Balkan Groups Flourish on the Net." Resonant Voices Initiative, May 5, 2017. <http://www.balkaninsight.com/en/article/far-right-balkan-groups-flourish-on-the-net-05-03-2017>
- Weedon, J., W. Nuland, and A. Stamos. "Information Operations and Facebook." Facebook, April 27, 2017. <https://fbnewsroomus.files.wordpress.com/2017/04/facebook-and-information-operations-v1.pdf>

---

<sup>i</sup> For the purposes of this paper, social media actors include Facebook, Instagram, YouTube, and Twitter, but the behaviors and trends named in this paper also affect other platforms.

<sup>ii</sup> Individuals interviewed from civil society and media organizations requested their names not be used in this paper.

<sup>iii</sup> These trends are exacerbated by the growing number of returned fighters from Syria and Iraq in the Western Balkans, as well as Indonesia and the Philippines.

<sup>iv</sup> Twitter noted at an open industry conference that behaviors are changing so quickly, understanding on the context and problem is critical, citing the use of bots to influence political discourse in the Philippines and Mexico. At the same conference, Juniper Downs noted Google regularly engages civil society academics and others to get input in an effort to better respond to changing behaviors and trends.

<sup>v</sup> Existing forums such as the Global Network Initiative, International Telecommunications Union, and Internet Governance Forum, say local organizations, have not provided the desired space and structure for constructive engagement on this issue.

<sup>vi</sup> In Macedonia, many groups focus on track false information spread around parliamentary elections to prevent violence or other activities that would have negatively affected the elections.

<sup>vii</sup> Rappler, a media organization in the Philippines, for example, regularly scrapes and collects data on political accounts to track bots, and other information operation campaigns around news events.

<sup>viii</sup> Based on interviews with leaders from the Philippines, Myanmar, Indonesia, Sri Lanka, Kosovo, Macedonia, Bosnia, Serbia, and Albania, there may be multiple reasons why this is the case.

<sup>ix</sup> This is part of a larger issue: an important entry point for major social media companies and user communities in other countries is through country governments due to the company's interest in operating and complying with the law in these countries.

<sup>x</sup> Conversations could build on existing discussions about information credibility measurement.

## Why We Feel Woefully Under-Informed About How We Can Do Our Jobs Better

Amy Sippitt, Full Fact

Full Fact is the UK's independent, non-partisan, factchecking charity. We provide free tools, advice, and information so that anyone can check the claims we hear from politicians, journalists, and other figures influencing public debate in the UK.

Simply publishing factchecks is not going to stop the spread of unsubstantiated claims or improve the overall accuracy of public debate. That idea has been discredited at least as far back as the 2008 US general election when FactCheck.org, one of the first generation of factchecking organizations, concluded that despite more aggressive fact-checking than ever before, “millions of voters were bamboozled anyway” (Jamieson and Jackson 2008). Research that does not take account of this is of limited use to us.

Full Fact is a second-generation factchecking organization. We see our dual role as ensuring reliable information is available, and stopping the spread of specific unsubstantiated claims. Africa Check and Argentina's Chequeado are two other examples. Publishing factchecks is just one tool in our work.

In the short term, we factcheck, which has three benefits: giving people reliable information to make up their own minds on big issues; scrutinizing the claims of people in public debate; and building an evidence base of how specific unsubstantiated claims arise and are spread.

Then, when we see specific unsubstantiated claims, we get those claims corrected at source. We've secured corrections from a Prime Minister and national newspapers, and we've worked with official statistics providers and charities to improve the way they present their information when it has been persistently misinterpreted.

In the medium term, we use the evidence from our factchecking to diagnose systemic problems and get systemic changes such as improvements to press complaints procedures.

In the long run, we are in a fight about the culture of public life.

We do not give claims accuracy ratings: we see our job as filling in shades of grey when campaigners often talk in black and white. We also use a variety of formats other than a written factcheck, such as a simple “claim and conclusion,” videos, graphics, as well as interviews on radio and TV.

We believe there is now a need for a third generation of factchecking: responding to challenges including (a) everyone having the power of access to public platforms, vastly increasing the range of monitoring needed; (b) pseudo-grassroots activity online up to state-sponsored disinformation; and (c) the declining relevance of existing standards processes.

Full Fact is running a number of initiatives to prepare for this. These include –

Organizing our work around debates not individual claims—to properly inform individuals’ *beliefs* we need to understand how individual claims on a particular topic build up to form a picture about the world.

Our world-leading automated factchecking work. The tools will first spot claims that have already been factchecked in new places, helping us monitor the spread of misinformation so we can target and evaluate our interventions effectively. Second, they will automatically detect and check new claims, freeing up our resources to focus on more complicated claims.

Our Third Generation Factchecking project which aims to develop, test, and implement effective reusable editorial packages that could appear anywhere from our website to search results.

### **Misinformation and disinformation in the UK**

The UK is a fortunate country with high levels of education, well-developed public and civil society institutions, and some highly trusted media. Nevertheless, there is evidence that the public is substantially misinformed on key issues of public debate. The Ipsos MORI Perils of Perception series—which asks participants to answer factual questions—has shown that for citizens in the UK and elsewhere across many different topics (Ipsos MORI 2016).

Observers have pointed to examples of inaccurate information in the UK:

“Politicians tend to promote the statistics that best present their case...In some cases, spinning reduces the story behind the statistics to such an extent that the picture is no longer true” (Jenkin 2013).

“The public debate is being poorly served by inconsistent, unqualified and, in some cases, misleading claims and counter-claims” (Treasury Select Committee 2016).

There is “sufficient evidence to conclude that some sections of the press have deliberately invented stories with no factual basis in order to satisfy the demands of a readership” with “sections of the press...prioritizing the worldview of a title over the accuracy of a story” (Leveson Inquiry 2012).

Our partnership with First Draft brought together verification and factchecking skills for online content during the 2017 general election and showed inaccurate information spreading enthusiastically online, from official, aligned, grassroots, and perhaps pseudo-grassroots sources. But this is not a pattern unique to digital spaces, nor is it new.

Full Fact leaves it up to our readers to judge where inaccuracies lie on the spectrum of misinformation (“the inadvertent sharing of false information”) and disinformation (“the deliberate creation and sharing of information known to be false”) (Wardle 2017). We see claims which seem to represent genuine misunderstandings as well as careless, reckless, and sometimes willful inaccuracies.

The most common causes we come across seem to us to be the careless use of information or a reasonable misunderstanding of a source. That is why well communicated, high quality, and easy to access information is hugely important to tackling misinformation and disinformation.

### **Research needs**

We welcome and are eager to act on research which –

Is not just US-based or is clearly transferrable to other countries. The political environment in the US is unique. For example, voters seem to have stronger partisan affiliations (Dalton 2016). In the UK, just 31% of the British public classify themselves as “very” or “fairly strong” supporters of a political party (Hansard Society 2017).

Tackles specific questions. Fake news is an unhelpful catch all term (Full Fact 2017).

Understands the range of interventions factcheckers globally use, from seeking corrections to public education work, and the range of existing and possible formats.

Has a sophisticated approach to change over time. Factchecking interventions cannot be seen individually, nor should their effect be viewed simply at a single point in time.

Clearly communicates its practical conclusions and lessons, including under what circumstances impacts have been found, how small or large the effect is, and the level of confidence in the findings.

Is methodologically rigorous and designed to be replicable, including publishing data.

As it stands we feel woefully under-informed from the research about how we can do our jobs better as much research in this field falls short of these hopes or is hard for busy factcheckers to access and apply.

We believe that there is great value in existing research from many fields. We are eager to see more practical and critical

literature reviews. In particular, we are keen to see cross-disciplinary research. The transmission of facts is not a new topic, and is not a research area unique to political scientists and psychologists. As well as pursuing new research, we hope to learn what academics in the fields of marketing, advertising, education, and elsewhere might be able to tell us.

### **Some specific research questions**

***What works to stop misinformation and disinformation spreading online and offline?*** How does information spread online and offline? Who should be held accountable? Are corrections worth the effort, especially when information spreads so quickly and easily online?

***How does factchecking contribute over time to people’s beliefs?*** The existing research has tended to focus on the impact of a participant reading one factcheck, and whether that changes someone’s beliefs. However, political campaigns take place over months, and beliefs form over years. The research needs to look at the impact of factchecking and associated activities over much longer periods if we are to properly understand its impact on beliefs.

***How can we communicate effectively?*** We know very little about what works to get a person to simply understand a factcheck. Do some formats work better than others? Do we need to tailor formats to individuals? How does context matter? We are doing some initial work on this with the University of Edinburgh’s Neuropolitics Research Lab to look at what works to communicate factual information in digital spaces.

***Who are we not reaching and how do we reach them? What does it take to secure trust?*** Are there people who are too

distrustful for our work? If so, how do we reach them?

***What interventions work to improve the skills of people receiving information?*** We know that individuals distrust but we also know that lots of individuals are misinformed. As Professor Onora O’Neill has said: “We may end up claiming not to trust, and yet for practical purposes place trust in the very sources we claim not to trust” (O’Neill 2002). We need to understand more about whether we need to make people more skeptical, so they are wary of what they see, or less cynical so they don’t feel the need to switch off from any factual messages. We also need to know how factchecking and verification influences individuals’ skepticism and cynicism. More specifically, there are journalist training programs and public education programs: what is their impact?

### **Increasing the cost of lying and the rewards for accuracy**

When the vast majority of the public says that they distrust politicians, is there any reason for politicians not to lie?

### **References**

Dalton, R.J. (2016) “Politics.” Oxford Research Encyclopedias.  
<http://politics.oxfordre.com/view/10.1093/acrefore/9780190228637.001.0001/acrefore-9780190228637-e-72#acrefore-9780190228637-e-72-figureGroup-1>

Hansard Society (2017). *Audit of Political Engagement*. London: Hansard Society.

House of Commons Treasury Committee, *The economic and financial costs and benefits of the UK’s EU*

*membership. First Report of Session 2016-17.* (2016)

Independence and Funding, Full Fact,  
<https://fullfact.org/about/funding/>

Ipsos MORI, (2016). “Perceptions Are Not Reality: What the World Gets Wrong.” Ipsos MORI,  
<https://www.ipsos.com/ipsos-mori/en-uk/perceptions-are-not-reality-what-world-gets-wrong>

Jamieson, K.H., and B. Jackson. (2008) “Our Disinformed Electorate,” FactCheck.Org,  
<http://www.factcheck.org/2008/12/our-disinformed-electorate/>

Jenkin, B. (2013) “PASC Demands Government Stats Are Presented With ‘The Whole Truth.’” Parliament.UK,  
<http://www.parliament.uk/business/committees/committees-a-z/commons-select/public-administration-select-committee/news/publication-of-communicating-statistics-report/>

Leveson Inquiry into the culture, practices and ethics of the British press. (2012). *Report of the Inquiry into the Culture, Practices and Ethics of the Press, Volume II*. London: The Stationery Office.

O’Neill, O. (2002) “Reith Lecture Four: Trust and Transparency.” BBC Radio 4.  
<http://www.bbc.co.uk/radio4/reith2002/lecture4.shtml>

Wardle, C. (2017) “Fake News. It’s Complicated.” First Draft,

<https://medium.com/1st-draft/fake-news-its-complicated-d0f773766c79>

Written Evidence Submitted by Full Fact (FNW0097). (2017). Parliament.UK, <http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/culture-media-and-sport-committee/fake-news/written/48260.html>

## Field Notes: Observation from a School Library

Mike Barker, Phillips Academy

### Welcome to our Wheelhouse

In months following the 2016 election the country became fixated on two new phrases: “Fake News” and “Post-Truth.” These two phrases were repeated and recycled across endless publications and news platforms, all the while presented as an emerging phenomenon we had never seen before. But, of course, we had. These terms simply described an information ecosystem plagued by disinformation--an historical, well-known threat to civil societies, and, in our case, a functioning democracy.

Since the election, we have experienced a scramble toward a solution primarily involving technologists, journalists, fact-checkers, media academics and, to an extent, educators. These collaborations have been fruitful and yielded a number of very creative and mostly technological solutions. But most solutions to date address the immediate problem of how disinformation circulates on our social platforms. In this way, we have started to think of this issue as more of a technological problem and as less one about knowledge and education. This distinction is important as it will lead us to more sustainable solutions going forward.

Recent evidence has surfaced suggesting agents with ties to the Russian government bought ad space on our most common technology platforms for the purpose of distributing disinformation (Shane and Goel 2017). For the short-term and current news cycle, this has given politicians scapegoats to blame. While it is clear these companies have work ahead of them, placing blame

there does not address the larger, more substantial problem that the impact of this disinformation campaign made evident.

As a country of information consumers, the past election exposed a dearth of critical thinking skills when engaging with text or visual information. It exposed a broad lack of understanding related to authority of a piece of information, and, even worse, lack of curiosity with regard to an information source and a lack of sophistication in addressing bias.

In the library profession, these skills are commonly referred to as Information Literacy skills. We are graduating students without these skills, for some reasons I will describe in more detail. These are skills school libraries and librarians are well-positioned to help develop in the next generation of information consumers. Thus, a more substantial, perhaps longer-term solution to the problem of a “Post-Truth” world is an approach that aims to integrate school library programs better with the rest of the curriculum. According to the American Library Association, we have close to 100,000 school libraries (Number 2015) already situated in our 135,000 public and private institutions (Fast 2017), and we would be smart to leverage the existing presence they hold within schools to better solve the problems we have. With the right partnerships, these libraries can bring significant strategic value to their collective learning communities: a broader advocacy network.

### **Advancing Inquiry as Pedagogy**

Positionally the library does not drive classroom activity; it supports it. No matter how effective a librarian may be in advocating for information literacy instruction, odds are that a librarian plays a secondary or even tertiary role to choices related to content and pedagogy. In more cases than not--perhaps due to many years of teaching to standardized tests--those choices usually follow a set curriculum, with a set of pre-selected texts. Even in the best case scenario, this approach usually includes assignments driven by predetermined prompts. Yet this flavor of teaching and learning does not really correlate with a world that is far more open-ended. It does not fully allow for a student to seek out information broadly or to practice assessing what they find. In a nutshell, this mode of teaching is response oriented, rather than inquiry oriented.

A shift to an inquiry oriented pedagogy would offer students opportunities to interact more deeply with a less curated information ecosystem. Short, sustained research topics where students formulate and investigate their own topics by interacting with the broad array of content on the web, as well as in libraries, would seem to be a more complementary and effective way to train information consumer habits in the future. We need a research agenda which investigates this notion of open inquiry as an effective pedagogical device. With better evidence on the benefits this shift could bring to learning, traffic would start to flow more heavily into school libraries.

### **Broadening the Scope and Shape of Information Literacy**

While libraries have long carried the the flag of information literacy, they are not free of blame either as it relates to the state of

things today. One responsibility libraries, and in particular school libraries, have been slower to assume is to look critically at the all of the skills needed today to be “information literate” across an array of mediums and source types. Information is communicated more than ever through visual content than ever before. In this way, instruction cannot be focused on simply assessing print content. How are we to prepare students to assess the authority of articles and tweets and other content produced by bots and artificial intelligent agents? This challenge will only get harder and more complex as AI gets more sophisticated. How are we preparing students with an awareness of the information ecosystems they inhabit and the ability to identify how pieces of information are digitally served to them? How are we preparing them with an awareness of their own biases and two-system cognitive tendencies as they interact with a piece of information?

Shaping a research agenda that might help us to broaden or modernize the traditional library’s sense of information literacy to something more holistic and reflective of the technological and cognitive aspects of the term is crucial. In doing so, we would help libraries better prepare information consumers in the future.

### **Expanding Partnerships**

An effective way to promote inquiry-based learning and to expand the definition of information literacy would be to shape partnerships with the very platforms the students use everyday. Google for Education has been incredibly successful selling products and technology to schools. Though this is a for-profit venture, it likely accelerated technology adoption in American classrooms. Yet, aside from the Google Books project, there really is no

analogous partnership Google offers with school libraries to better develop information literacy in students. While it would be wise not to partner too closely with one group that monetizes information and data, a collaboration of this kind could be beneficial to both sides. A Google-led curriculum on information literacy is the kind of partnership that would attract the attention of teachers and drive activity back toward school libraries.

### **Libraries: Defunded & Underutilized ... We Need New Allies**

If implemented, strategies like some of the ones listed above would augment the value proposition of school libraries for the communities they serve, and help reverse a lengthy trend of divestment. This divestment is usually couched in the view that school libraries are underutilized because “we have Google now;” but this is precisely why they matter more than ever before.

For the most part, the library community has relied on a strategy of outreach and advocacy, but library advocacy cannot be the only solution--and to that end those advocating for the value of libraries cannot primarily be people working within them.

Jim Neal, the incoming President of the American Library Association (ALA) recently stated to *School Library Journal* that his focus would be on school libraries.

“With weakened school libraries, we create a nation that is not able to locate and evaluate information that is accurate and balanced. Knowledge literacy is fundamental,” he said (Bayliss 2017).

Libraries were initially established as a means to ensure that citizens operating in a functioning democracy would have access to the information they needed to participate in

civic life. In much the same way, Neal’s diagnosis signals the role school libraries can play in supporting and strengthening democracy in America. But to realize this mission we need to catalyze programmatic change within the library, and to increase support for these programs with tested and valid research.

### **Resources**

Bayliss, Sarah. “Incoming ALA President Jim Neal: Supporting School Libraries.” *School Library Journal*, May 24th, 2017.  
<http://www.slj.com/2017/05/industry-news/incoming-ala-president-jim-neal-supporting-school-libraries/#>

“Fast Facts: Educational Institutions,” National Center for Education Statistics, Accessed October 21, 2017.  
<https://nces.ed.gov/fastfacts/display.asp?id=84>

“Number of Libraries In the United States,” American Library Association, Last updated September, 2015.  
[http://www.chicagomanualofstyle.org/tools\\_citationguide/citation-guide-1.html](http://www.chicagomanualofstyle.org/tools_citationguide/citation-guide-1.html)

Shane, Scott and Vindu Goel. “Fake Russian Facebook Accounts Bought \$100,000 in Political Ads,” *New York Times*, September 6, 2017.  
[https://www.nytimes.com/2017/09/06/technology/facebook-russian-political-ads.html?\\_r=0](https://www.nytimes.com/2017/09/06/technology/facebook-russian-political-ads.html?_r=0)

## How Academics Can Help Platforms Tackle Disinformation

Nic Dias, First Draft

It was over a year ago that Craig Silverman published the report that brought wide attention to the mis- and disinformation space (Silverman 2016). Journalists and fact-checkers have since reported on countless rumors and hoaxes, and social media researchers have used what data they have to track attempts to amplify this information.

Google, Facebook, and Twitter have finally begun to acknowledge the scale of the mis- and disinformation problem on their platforms. But they continue to fall short of any comprehensive response or transparency. This pace is frustrating for many disinformation scholars. Yet little has been written about what research institutions might do to help these technology companies respond to issues surrounding the quality of information shared on, or surfaced by, their platforms.

In the hope of better understanding how disinformation researchers have interacted – and might better interact – with platforms, I spoke on background with platform representatives and academics about their experiences working with one another. Notably, these interviews pointed to interlocking, if not similar, sets of problems disrupting cooperation. And a majority even brought up similar ideas about how to address those problems.

It's important to acknowledge the potential conflicts that arise when academics begin to play a role that could reasonably be compared to R&D. To be clear, I'm not advocating that researchers devote

themselves to responding to needs of the technology companies. Rather, I provide evidence that thoughtful cooperation between research institutions and platforms would benefit the development of effective, short-term responses to disinformation, and so must be a priority.

### **Some research institutions can act as honest brokers between stakeholders.**

The benefits of bringing together journalists, researchers, and platforms are myriad. For one, institutional and political factors can make it difficult for platforms to explore some research questions. One interviewee explained that research within industry is often overly constrained by the financial calendar and competition between employees, particularly in companies where the bottommost performers are regularly replaced. As a result, emphasis is placed on projects that can produce results within the quarter, even when such projects may have negative long-term side effects.

Cooperation between stakeholders would also allow them to fully benefit from each other's expertise, and to pool resources in incredibly powerful ways, like cross-platform databases. As one platform representative described:

"I have visibility into a slice of the challenge that... our partners touch. If we're thinking about really trying to solve, or at least seriously make a dent in this misinformation challenge, that also touches [Platform X], that also touches [Platform Y]."

Of course, competitive frictions within each sector, and broad mistrust between them, complicate collaboration. There's also an undeniable mismatch between technology companies' move-fast-and-break-things approach and academia's measured methods of progress.

Yet several platform representatives suggested that some research institutions – like academic institutes or more practically-oriented labs – are best positioned to act as trusted brokers between journalists and technology companies. For one, they are relatively trusted by all parties, including the public, particularly where they are charitably funded. As one platform representative explained:

"Think tanks have essentially denigrated into lobbying shops, which has created the need for some sort of neutral hub for projects that require complex coalitions... Working with academic institutes creates the ability to move a project with a reasonable velocity, in a way where there's that layer of intermediation between [platforms] and players that, for whatever set of reasons, it might be hard for us to directly build a relationship with."

Academic institutes also commonly employ people who have worked in industry. This is an advantage for two reasons. First, ex-industry has a better sense how to work with industry. Second, those partnerships which have succeeded were built on personal relationships between current and former employees or students. One academic interviewee said the following:

"I've had personal experiences where I thought that I had great ideas but I didn't have the connections, and so they weren't pursued. And then I had some other cases in which I had ideas that were probably kind of

mediocre, but they were pursued just because I had the network contacts."

Research institutions can build on this social currency by valuing industry experience when hiring employees, or valuing industry ambitions when selecting student interns and assistants.

### **Be more solution-oriented researchers.**

Another theme from my interviews was that platform representatives struggle to apply research when it abstracts away from particular platforms. They were also discouraged with what they saw as research's tendency to complicate questions rather than answer them:

"That work is actually 95% useless to industry. Because we know that it's hard. We know that it's complicated. We know that it's ambiguous. But we've got to do something. So, reading an analysis of why it's hard doesn't move us forward."

Ultimately, what product managers need is something to convince engineers to prioritize a feature. Thus, platform representatives said it would be helpful if researchers crafted their research, or at least their discussion sections, with an eye to the circumstances in which platforms operate. Specifically, that means providing explicit suggestions to the platforms where possible, rather than assuming a study's implications are clear. It also means considering how one's research might apply internationally, and being more solution-minded and platform-specific. As one platform representative said:

"If it's Google-specific, or YouTube-specific, or Facebook-specific – it's generally more useful because we can understand more concrete conclusions based on the actual products being used, as

opposed to more general statements about here's how people consume information..."

Platform representatives also remarked that it had become impossible to keep up with all the new disinformation studies being published. Indeed, the recent influx of interest and funding for disinformation research has been accompanied by a still-swelling flood of research. Rasmus Klein Nielsen (2017), for example, found that 17 percent of the papers presented at the most recent Future of Journalism Conference at Cardiff University were focused on "post-truth, truth, and fake news."

As such, platform representatives said brevity in research write-ups is greatly appreciated. One researcher also suggested that more papers reconciling and translating the immense volume of at times contradictory research would be beneficial. These papers could even be aggregated in a cross-disciplinary journal of disinformation.

### **Consolidate conferences to minimize time demands.**

Another product of the sudden focus on disinformation is an overabundance of conferences organized about the topic. One platform representative described it this way:

"It gets to a point where every other day there is a different conference on the same topic. And it kind of becomes like a traveling carnival, where the same people just move around different hotels and different conference venues."

Unfortunately, these conferences tend to devote most of their agendas to the presentation of papers. This, several platform representatives suggested, is not an optimal use of everyone's time, brainpower, and experience. It's a waste to expend the

effort to bring together such an exceptional group of people, only to have them to sit in an audience, listening to content they could have already read – and likely have. As one representative put it:

"It's not always the best use of time to listen to a panel of four academics, who have all published papers that we've all read, regurgitate what's in their papers... Conferences should be about discussing what your paper says – the wider framework, the policy recommendations."

The same platform representative suggested trading the current conference schedule for one with fewer, longer gatherings. This would ensure several things: It would minimize time spent traveling and away from the office. It would save everyone's money. It would cut down on redundancies between conferences. And it would ensure everyone can meet.

To make the most of these gatherings, it may also be necessary to provide space for platform representatives to talk under Chatham House rule, or even off-the-record. This would provide a genuine opportunity for the powerful players collected to combine their perspectives and build trust:

"If an event is on the record or quotable it makes it really, really hard for companies to discuss hypotheticals or potential policy solutions before we have made decisions internally. Because, ultimately, we work for publicly traded companies that attract a large amount of public interest... And one errant phrase on the record in an event that gets reported can quickly turn into a news story that doesn't reflect the conversation and context, and [is] blown out of proportion."

## Conclusion

The centrality of social media and search engines, and the apparent scale of the information pollution on them, suggests that platforms will play an important role in relegating disinformation to the depths of the internet – at least in the short-term. However, as broadly trusted and independent experts, academics appear to be well-positioned to strengthen their efforts by informing and convening stakeholders.

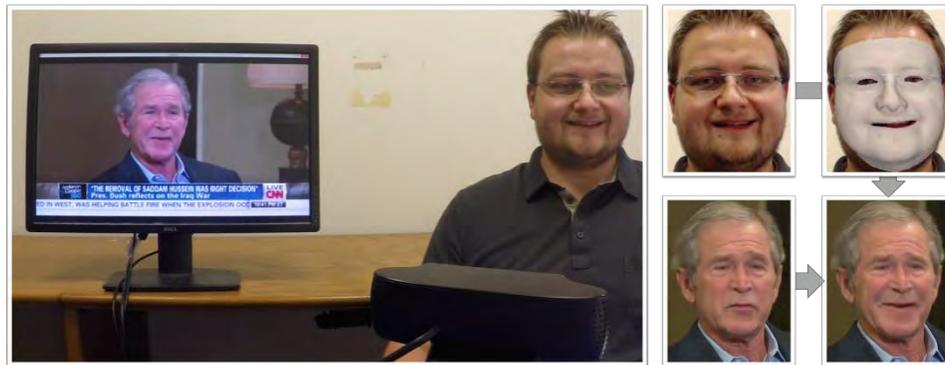
As I said at the top of this paper, many scholars may be hesitant to work with industry. Yet none of my interviewees suggested that academics should stop seeking truth for truth's sake. Rather, they argue that a willingness to cooperate with platforms would go a long way to help them tackle disinformation.

## References

- Silverman, C. (2016). This Analysis Shows How Viral Fake Election News Stories Outperformed Real News On Facebook, BuzzFeed, <https://www.buzzfeed.com/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook>
- Nielsen, R.K. (2017). What is Journalism Studies Studying? [Blog post]. Retrieved from <https://rasmuskleisnielsen.net/2017/09/14/what-is-journalism-studies-studying/>

# Face2Face: Real-time Face Capture and Reenactment of RGB Videos

**Justus Thies**, Technical University of Munich; University of Erlangen-Nuremberg  
**Michael Zollhöfer**, Max-Planck Institute for Informatics  
**Marc Stamminger**, University of Erlangen-Nuremberg  
**Christian Theobalt**, Max-Planck Institute for Informatics  
**Matthias Nießner**, Technical University of Munich



**Abstract:** Face2Face is an approach for real-time facial reenactment of monocular target video sequences from the Internet. The goal of Face2Face is to animate the facial expressions of the target video in real time based on the input of a commodity webcam that captures the source actor. To this end, Face2Face tackles a hard, inverse rendering problem. It reconstructs the facial geometry of both faces and tracks the non-rigid motions. Using a novel expression transfer technique, the facial expressions are transferred from the recovered source face to the target face. Finally, a convincingly re-rendered synthetic video stream is produced based on novel image based rendering techniques.

## Introduction

Our primary goal is to create a mathematical model of our world. Such a model enables computers to reconstruct, understand, and interact with it. Computer Vision tries to obtain this model from image data. The reconstruction of the surrounding is very important nowadays. New technologies like Virtual Reality or Augmented Reality rely on such data. The better the quality of the representation is, the better these new technologies will work.

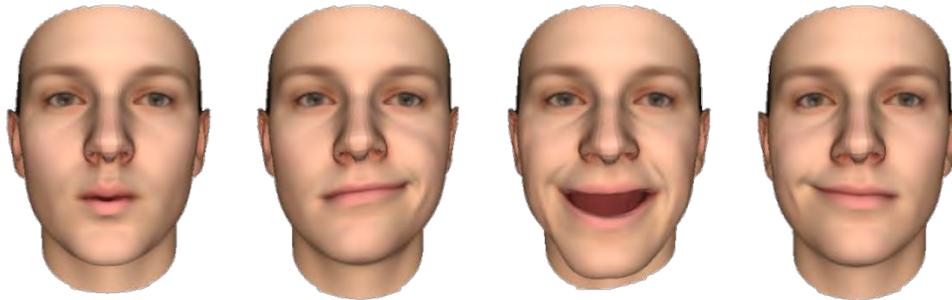
In contrast of scanning solid objects with e.g. a Kinect depth sensor, Face2Face concentrates on the harder problem of reconstructing moving faces in uncontrolled environments. Most existing real-time face trackers are based on sparse features and thus capture only a coarse face model. Face2Face tries to use all available information in the captured input, i.e., every pixel, which is why it is called a dense face tracker. The method follows the principle of analysis-by-synthesis, which means that we determine the parameters of our face model

by minimizing the error between the input image and a synthesized face image. Our resulting synthesized model is so close to the input that it is hard to distinguish between the synthesized and the real face captured by the RGB camera.

### Method Overview

As described above the primary aim of Face2Face is to reconstruct and track a mathematical model of the human face. To model geometry and skin appearance, Face2Face is based upon a database of 200 scanned faces and 76 modelled expressions. Thus, using this data new faces can be generated using a relatively small set of parameters.

Figure 1: Synthesized faces with different expressions using Face2Face



Using the standard graphics rendering pipeline of modern computers, we can generate synthetic images as shown in Figure 1. This is key for the analysis-by-synthesis approach that we apply to reconstruct the face of a person in a video stream.

The analysis-by-synthesis approach is an iterative process of synthesizing and error minimization. The error in our method describes the difference between an input image of a real camera and the synthetic image of the currently reconstructed face.

Face2Face uses a dense error metric thus comparing every single pixel of the two images. This requires a lot of computation that is only feasible to be done on modern Graphics Cards (GPUs). The optimizer that is used to minimize the error is a highly optimized parallel GPU-based Gauss-Newton solver that enables real-time framerates.

In Figure 2, we show a sequence of reconstructed faces.

Figure 2: Facial motion capturing based on the principal of Analysis-by-Synthesis: Top row shows the input and the bottom row shows the recovered facial geometry.



Based on this estimation of a face we implemented facial reenactment.

To this end, we reconstruct the face of both the source actor and the target actor using the approach described above. The source actor is tracked in a live video stream, while the target video is preprocessed. Using the variation of mouth motions in the target video, we create a database of mouth appearances, especially the appearance of the mouth interior.

During runtime we then transform the geometry of the target face to match the reconstructed expression of the source actor. Using the deformed face geometry of the target face and the mouth appearance in the database that is closest to the applied deformation we re-render the face on top of the target video.

Figure 3 demonstrates the effectiveness of the described approach.

Figure 3: Facial reenactment applied to a video from the Internet.



### Use Cases

There are many possible use cases for our dense face tracking technique. The most

prominent use case is the usage in post-production systems in the film industry. There are already techniques that transfer

the expressions of a human to a virtual avatar. This is also possible with our technique, despite the fact that we can even transfer the expressions / mouth movement to another human actor without the need of special hardware and long scanning procedures. Thus, we can easily alter the face of an actor after capturing. E.g., one can alter the lighting, the skin or modify the expression. We also think that our technique paves the way to live dubbing in a teleconferencing scenario. A similar field where such a dense face tracking could be beneficial, is the game industry. Here, the motion of a player can be transferred to a virtual avatar of the same person in an online multiplayer game, resulting in a more realistic output.

Our project stems originally from another important field - medical research. There we analyzed the healing of patients that suffer from a cleft lip and palate disorder. Head tracking is also important for other medical purposes, i.e., tracking during surgeries etc. Aside from that, our technique can be adapted to track other body parts or organs. Which is one of the most challenging tasks in modern medicine and is for example needed to damage as less as possible healthy matter during a therapy. Also, many psychologists are interested in our system to be used to treat people that suffer from psychical diseases. There is also social psychological research that analysis the bias of which person is more trustworthy.

Fraud detection: As we reconstruct the face parameters as well as the lighting conditions, our approach can be used to

detect inconsistencies. For example, the expressions of a person and also the transition between them are unique. Thus, a fraud can be detected by analyzing the tracked expressions in a video sequence and comparing them to a reference video sequence. Which is similar to the analysis of handwritten text (graphoanalysis). In the field of Man-Machine Interaction detection and tracking of movements is the key-component. Thus, accurate tracking pushes the development of these applications.

### **Discussion and Conclusion**

The demonstrations of Face2Face at several conferences and in the supplemental video on Youtube resulted in a controversial discussion of the developed technique. People, who watched our system for the first time, were thrilled by the results of our method and thought of funny things they could do with it. But soon they realized that it could also be used to manipulate videos for propaganda or other evil purposes. In general, they are right with this rating, but it is important to sensitize people to such video manipulations. Existing methods that are used for movie production bring dead actors virtually to life, but nobody thinks of the abuse of such a technology. Demonstrating the simplicity of video manipulation teach the people to not blindly trust videos from unknown sources.

Nevertheless, we believe that this system paves the way for many new and exciting applications in the fields of VR/AR, teleconferencing, or on-the-fly dubbing of videos with translated audio.