2021 PUBLIC-PRIVATE ANALYTIC EXCHANGE PROGRAM

Combatting Targeted Disinformation Campaigns

A whole-of-society issue

Part Two

August 2021

2021 PUBLIC-PRIVATE ANALYTIC EXCHANGE PROGRAM

Combatting Targeted Disinformation Campaigns

A whole-of-society issue

Part Two

August 2021

United States

DISCLAIMER STATEMENT: This document is provided for educational and informational purposes only. The views and opinions expressed in this document do not necessarily state or reflect those of the United States Government or the Public-Private Analytic Exchange Program, and they may not be used for advertising or product endorsement purposes. All judgments and assessments are solely based on unclassified sources and are the product of joint public and private sector efforts.

Acknowledgments

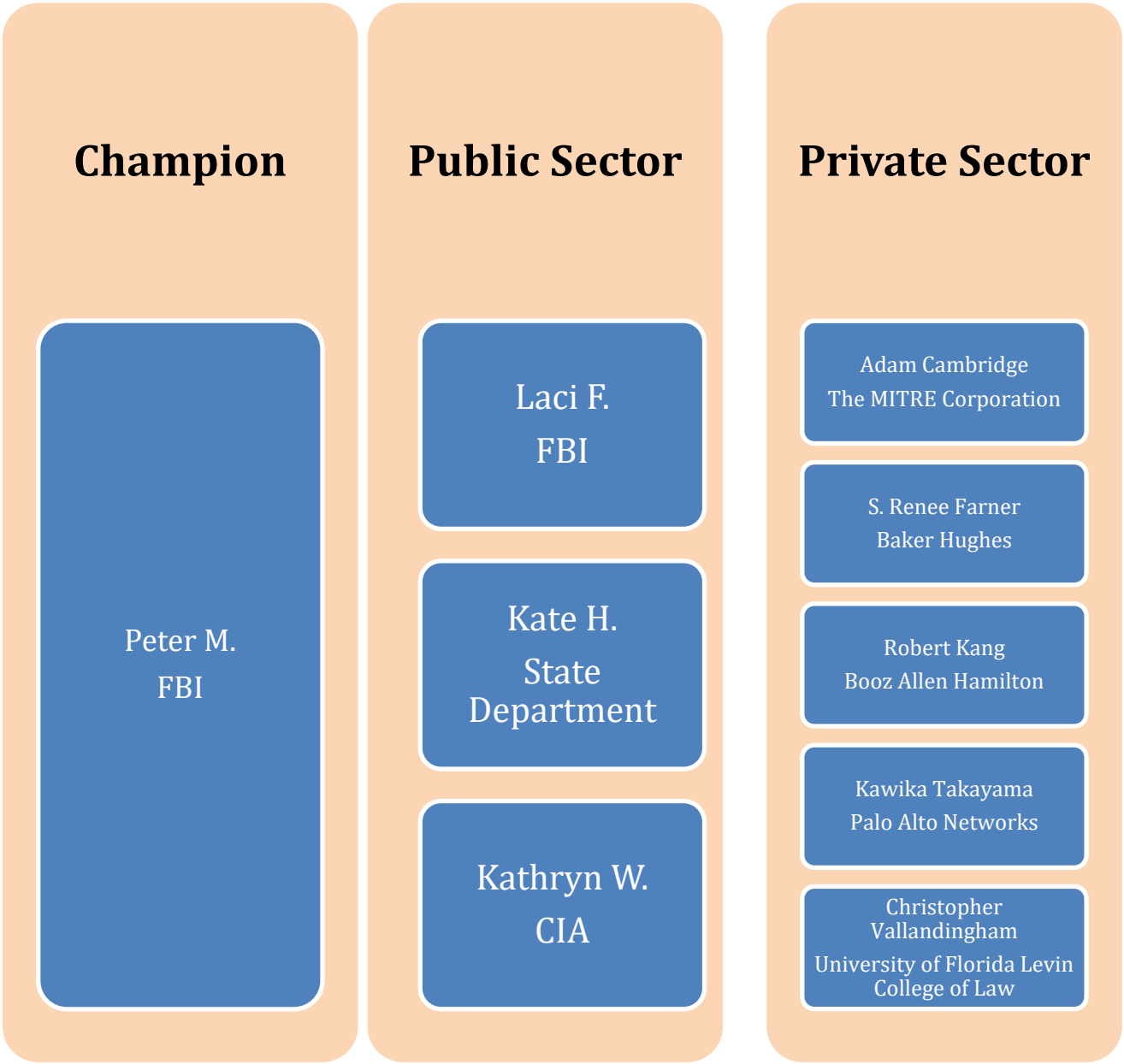
We would like to thank all who contributed to this research amid the unprecedented challenge of the COVID-19 pandemic. This project would not be possible without the assistance of the Office of the Director of National Intelligence and the Department of Homeland Security, who provided us the opportunity to work together on a very important and pressing topic.

We are grateful to the individuals from academia, private sector companies, and think tanks who joined us in our weekly teleconferences to lend their insights on the vexing topic of disinformation campaigns. Their input provided invaluable guidance on lines of inquiry and potential criteria for evaluating the effectiveness of different approaches to countering disinformation campaigns.

We would like to extend a special thank you to Peter M., Team Champion, whose patience and wisdom help guidance us through both years of our project.

We would like to acknowledge the Team members listed below for their contributions on this project and report. We are deeply indebted to Peter Thielen from Booz Allen Hamilton for helping the team to develop a graphic on how to spot dis/misinformation and bringing our vision to fruition and Hale Vallandingham for her assistance with graphics and formatting.

Team Members



Scope

This paper was produced by the Combatting Targeted Disinformation Campaigns team under the auspices of the Public-Private Analytic Exchange Program – an initiative of the Office of Director of National Intelligence and managed by the Department of Homeland Security. The paper was based on open-source research and interviews with subject matter experts. All judgments and assessments are based solely on unclassified sources, are the product of joint private sector and U.S. government efforts, and do not necessarily represent the judgments and assessments of the employers of the Team members.

Offensive and defensive cyber operations conducted by the U.S. government against foreign threat actors are beyond the scope of this paper.

Executive Summary

Recent events have demonstrated that targeted disinformation campaigns can have consequences that impact the lives and safety of information consumers. On social media platforms and in messaging apps, disinformation spread like a virus, infecting information consumers with contempt for democratic norms and intolerance of the views and actions of others. These events have highlighted the deep political and social divisions within the United States. Disinformation helped to ignite long-simmering anger, frustration, and resentment, resulting, at times, in acts of violence and other unlawful behavior.

Figure 1 Disinformation (obtained from Defense.Info).

All information consumers are vulnerable to being deceived by imposters, charlatans, hucksters, con men, and self-proclaimed experts. But ideological rigidity and intolerance of opposing views make information consumers especially vulnerable to such deception. In polarized environments, threat actors find ample opportunity to spread disinformation. Their voices are amplified by disgruntled audiences willing and sometimes eager to spread messages of discord.

In 2019, the Combatting Targeted Disinformation Campaigns team submitted the first of a two-part report on targeted disinformation campaigns.[1] In the first report, we provided

[1] 2019 Public-Private Analytic Exchange Program, Department of Homeland Security, *Combatting Targeted Disinformation Campaigns: A whole-of-society issue*, PDF file, October 2019, https://www.dhs.gov/sites/default/files/publications/ia/ia_combatting-targeted-disinformation-campaigns.pdf (Accessed August 24, 2021).

a broad overview of targeted disinformation campaigns, described how disinformation enters the information ecosystem, how threat actors exploit modern technology, and how information consumers wittingly or unwittingly contribute to the spread of disinformation. We then suggested ways to counter these campaigns. We concluded that, to combat these campaigns, a comprehensive solution involving many sectors of society and lines of efforts was required.

In our second paper, we expand on two themes explored in the first paper: 1) how to stem the supply of disinformation; and 2) how to reduce the demand for disinformation. In this paper, we recommend approaches that impact both supply and demand. None of these approaches are new; nor are they decisive by themselves. When combined and implemented consistently, the sum is greater than the parts.

Figure 2 Dilok Klaisataporn, Shutterstock.

Disinformation should not be viewed as a problem to be solved, but as a condition to be treated. There is no cure. However, preventative and alleviatory measures can be taken.

Recommendations

Source attribution	•Identification of sources of disinformation, when feasible. We believe that threat actors should not be permitted to hide behind the veil of anonymity and identifying them by name gives information consumers important information as they evaluate the truthfulness of the information.
Exposure of imposters	•Exposure of imposters who use false persona and fake credentials to dupe information consumers online. If possible, we believe that uncovering imposters should be done without revealing personally identifiable information about the person.
User control over content	•Greater control by information consumers over the sources of information that appear in their content feeds on social media platforms. We believe that giving information consumers greater control may help to limit the creation of echo chambers.
Fact checking	•Access to content alerts which identify information that is factually incorrect, fraudulent, misleading, or satirical. We believe that such alerts should be politically neutral and based on clear definitions of the categories used.
Psychology of disinformation	•Making information consumers aware of how they process online information and how their online activities facilitate disinformation campaigns will help them make better decisions regarding this information.
Media literacy	•Information consumers who understand how to evaluate the impact of social media and other modern forms of communication on their ability to assess the trustworthiness of information will be more resilient to disinformation.

Table of Contents

Figures

1. Disinformation Campaigns Are National Security Threats

Disinformation campaigns are threats to national security because they have a corrosive impact on democratic institutions and civil society in the United States. A healthy democracy depends on well-informed citizens, the competition of ideas, and the willingness to compromise. Disinformation campaigns undermine all three.

Disinformation campaigns have contributed measurably to divisions within U.S. society. Threat actors take advantage of these divisions and harness the power of social media platforms to spread disinformation to large audiences. At times, these disinformation campaigns influence the real-world behavior of information consumers. For example, disinformation about the 2020 presidential elections impacted the lead-up to the events of January 6, 2021 in Washington, D.C. Furthermore, disinformation about the COVID-19 vaccination program has undermined efforts to curb the virus and its variants. Disinformation also weakens international alliances and undermines U.S. attempts to project soft power abroad.

Figure 3 Jimmy Margulies via AP (from ABC news).

Figure 4 Divided country (obtained from News Bharati).

The launch of Sputnik I in 1957 sent shockwaves through the U.S. government. Fearing that the Soviet Union was surpassing the United States in technological and military prowess, the U.S. government responded by investing heavily in human capital and the development of technology. Experts identified weaknesses in educational institutions from primary to graduate levels, insufficient investment in basic and applied research, and governmental bureaucracies ill-adapted to guide the changes deemed necessary to respond to the threat. Major sectors of society contributed to a whole-of-society response to the perceived threat. A similar response is needed today for disinformation campaigns. However, the divisions in our society hinder such a coordinated response.

> *Perhaps because the disinformation crisis has descended upon us so suddenly, and because it reinforces our increasing political polarization, we've tended to regard it as inevitable and unavoidable—a fact of digital life. But we do have options, and if we come together to exercise them, we could make a meaningful difference.*
>
> *Glenn Gerstell, former NSA General Counsel*

In Finland, information literacy forms part of the national curriculum for primary and secondary schools. In math courses, instructors show students how statistics can be manipulated for the purposes of misleading others. In art classes, students learn how images can be altered to change their meaning. In language classes, students are exposed to different ways words can be used to mislead. And finally, students learn in history classes about past propaganda campaigns.

In Malaysia, the Communication and Multimedia Ministry launched an information verification portal called "sebenarnya.my." The public can use this portal to request that the Ministry verify information found online.

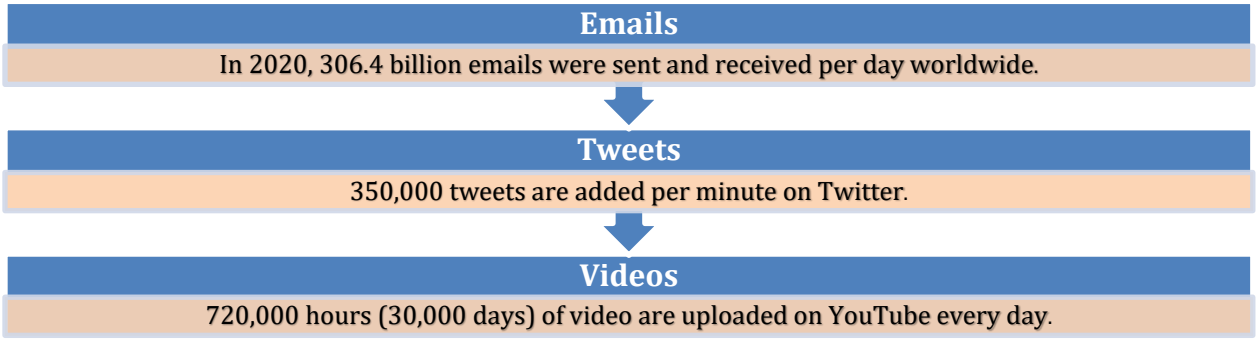
Germany introduced the "Act to Improve Enforcement of the Law in Social Networks" in 2017. This bill requires social media companies with over two million users in Germany to enforce 21 statutes in the German criminal code related to hate speech on their platforms. Platforms must review and delete unlawful content within 24 hours of receiving a complaint or be fined up to 50 million EUR.

Efforts of other countries: Finland [2]; Malaysia[3]; Germany[4]

[2] Jon Henley, "How Finland Starts Its Fight against Fake News in Primary Schools," *The Guardian*, last modified January 29, 2020, https://www.theguardian.com/world/2020/jan/28/fact-from-fiction-finlands-new-lessons-in-combating-fake-news (accessed August 24, 2021).
[3] Fairuz Mohd Shahar, "Communications Ministry launches sebenarnya.my to quash fake news, information", New Straits Times, last modified March 14, 2017, https://www.nst.com.my/news/2017/03/220604/communications-ministry-launches-sebenarnyamy-quash-fake-news-information (accessed August 24, 2021).
[4] Nina Jankowicz and Shannon Pierson, "Freedom and Fakes: A Comparative Exploration of Countering Disinformation and Protecting Free Expression", PDF file, Wilson Center, December 2020, https://acrosskarman.wilsoncenter.org/sites/default/files/media/uploads/documents/WWICS%20STIP%20Freedom%20and%20Fakes%20A%20Comparative%20Exploration%20of%20Countering%20Disinformation%20and%20Protecting%20Free%20Expression.pdf (accessed August 24, 2021).

2. Reducing the Supply of Disinformation

Methods exist which can help slow, but not eliminate, the supply of disinformation. The challenge is breathtaking. The global information ecosystem enables the creation and distribution of information on an enormous scale.



The global dissemination of information: Emails; [5] Tweets;[6] YouTube.[7]

In this section, we conclude that technological and non-technological solutions exist that can reduce the likelihood that disinformation will course through the information ecosystem and negatively impact the target audience. These methods include source attribution, fact-checking, and greater user control over algorithms used by social media platforms.

[5]Joseph Johnson, "Number of sent and received e-mails per day worldwide from 2017 to 2025", last modified April 7, 2021, https://www.statista.com/statistics/456500/daily-number-of-e-mails-worldwide (accessed August 24, 2021).
[6] "Twitter Usage Statistics," https://www.internetlivestats.com/twitter-statistics. (accessed August 24, 2021).
[7] Maryam Mohsin, "10 YouTube Stats Every Marketer Should Know in 2021" (Infographic), last modified January 25, 2021, https://www.oberlo.com/blog/youtube-statistics (accessed August 24, 2021).

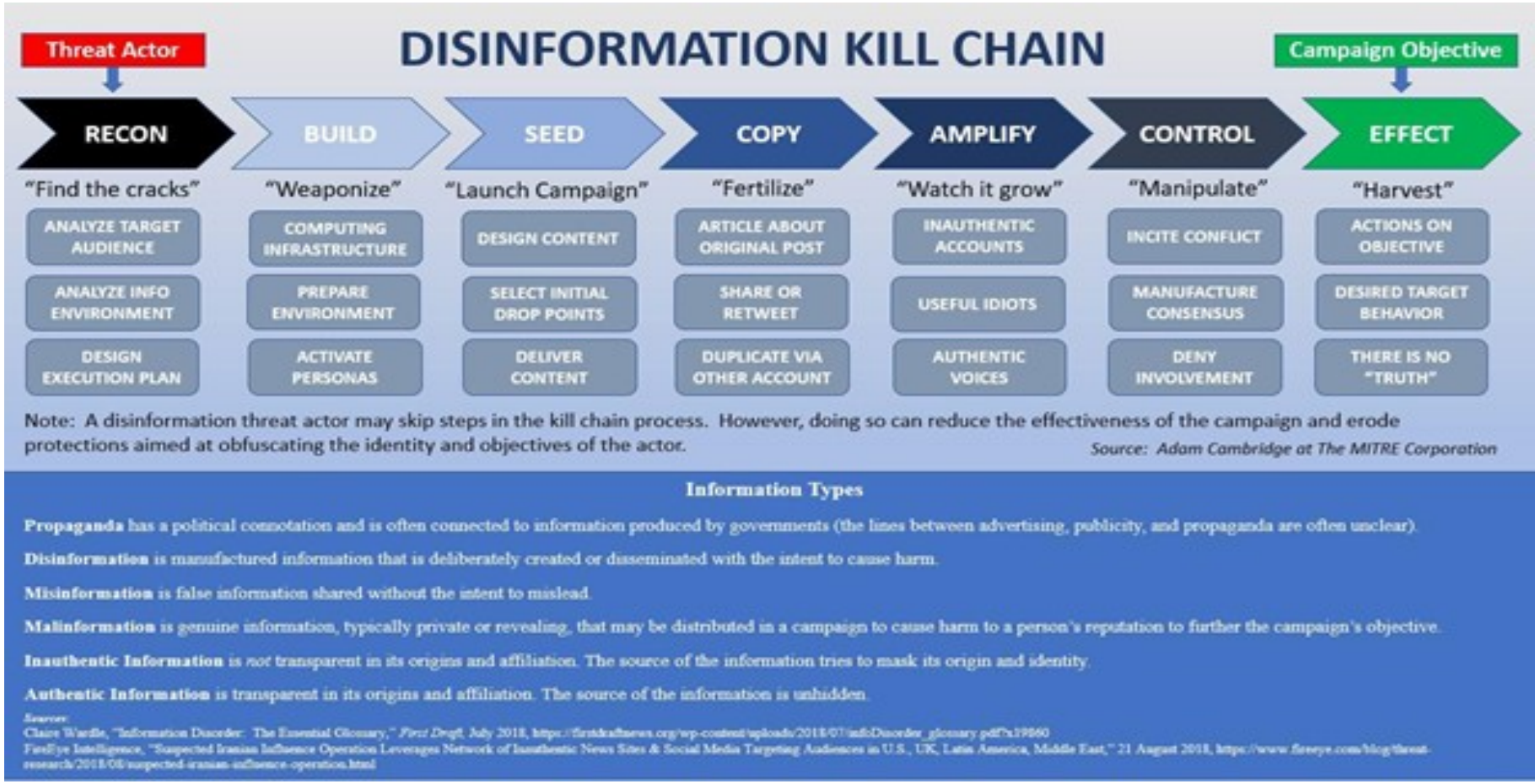


Figure 5 Source: The MITRE Corporation.

2.1. Source Attribution and Anonymity

Providing information about the threat actors responsible for disinformation campaigns can help slow the spread of disinformation since the source of information influences how information consumers evaluate the truthfulness of the information and the decision to later share the information. This method is called source attribution. Source attribution can be conducted in a manner that safeguards the constitutional and statutory rights of U.S. citizens to privacy and the freedom of speech.

In the United States, anonymity has long been considered a necessary component of the freedom of speech. However, anonymity on the Internet is a mixed bag. It not only promotes the discussion of sensitive topics, but also facilitates uncivil behavior and the spread of disinformation. In the next two sections, we consider source attribution of both foreign and domestic threat actors.

2.1.1. Identifying Foreign Actors

"Naming and shaming" is an approach to countering disinformation in which threat actors behind disinformation campaigns are publicly identified. When threat actors use fake personas to deceive, information consumers may decide to disregard the threat actor and the disinformation associated with that threat actor when the deception is uncovered. Once a threat actor is publicly identified, they will be shamed into altering course.

> Our deeply-felt national scruples about misidentifying a fake account or inadvertently silencing someone, however briefly, create a welcoming environment for malign groups who masquerade as Americans or who game algorithms….When tech platforms or regulators strive to take meaningful action to suppress abuse of their platforms and our American polity, there are waves of outrage over censorship. We have conversations about whether or not bots have the right to free speech, we respect the privacy of fake people..
>
> New Knowledge, ***The Tactics & Tropes of the Internet Research Agency***

Private companies have publicly attributed malicious cyber incidents to foreign state actors.[8] The benefits of doing so are usually short-lived since the threat actor will likely change tactics in response. These companies have stated several reasons for engaging in this activity, including supporting internal government discussions by allowing employees not possessing the appropriate security clearances to view information that would likely be classified if provided by the federal government, providing corroborating information for other information gathered by the federal government, and underscoring to the public the reach of the cyber incident.[9]

One method of disseminating this information is through source alerts. Researchers at Harvard University tested whether source alerts help to reduce the likelihood that

[8] Sasha Romanosky and Benjamin Boudreaux, "Private Sector Attribution of Cyber Incidents: Benefits and Risks to the U.S. Government", PDF file, Rand, January 2019, https://www.cs.dartmouth.edu/~ccpalmer/teaching/cs55/Resources/Papers/RAND_WR1267.pdf (accessed August 24, 2021).
[9] Ibid., 17-20.

information consumers will believe or share political messages. Limiting the study to foreign sources of information and two social media platforms (Twitter and Facebook), the researchers explored whether general ("foreign government") and specific ("Russian government") source alerts did have such an effect.[10] In assessing how the participants in the study viewed the truthfulness of the information, the researchers found that the effects were more significant for specific source alerts than for general ones. Exposure to specific source alerts reduced the probability that the participants would find the disinformation truthful. General and specific source alerts reduced the tendency to "like" or share the disinformation on Twitter only.

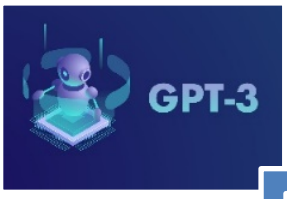
Figure 6 State-affiliated media account labels (obtained from Twitter)

FireEye, CrowdStrike, Dell SecureWorks, and Cisco Talos have publicly identified foreign threat actors.[11] In recent years, both Facebook and Twitter have increased the use of specific source alerts and have tied disinformation campaigns to foreign actors such as the Iran Broadcasting Company and the Royal Thai Military.[12]

[10] Jason Roos Arnold et al., "Source alerts can reduce the harms of foreign disinformation", PDF file, *Harvard Kennedy School Misinformation Review*, May 2021, https://misinforeview.hks.harvard.edu/wp-content/uploads/2021/05/arnold_source_alerts_foreign_disinformation_20210510.pdf (accessed August 24, 2021).

[11] Romanosky and Boudreaux, "Private Sector Attribution of Cyber Incidents: Benefits and Risks to the U.S. Government." 6.

[12] Josh A. Goldstein and Shelby Grossman, "How disinformation evolved in 2020", Tech Stream, Brookings, last modified January 4, 2021, https://www.brookings.edu/techstream/how-disinformation-evolved-in-2020 (accessed August 24, 2021).



Developed by OpenAI, GPT-3 uses neural networks and machine learning to generate automated text in response to prompts from humans. These texts are difficult to differentiate from those written by humans.

Figure 7 OpenAI, GPT-3.

There are risks associated with source attribution. Such source attribution invites blowback from these threat actors and may undermine ongoing federal law enforcement, intelligence, and diplomatic efforts.

The evidence that such naming "shames" foreign actors into modifying their behavior is thin.[13] However, identifying specific foreign threat actors behind disinformation campaigns gives information consumers the opportunity to evaluate the information with this source attribution in mind and therefore we recommend such attribution when circumstances allow.

❖ Recommendation 1: Identify Foreign Sources of Disinformation

2.1.2. Identifying Domestic Actors

Since the 2016 U.S. presidential election and the emergence of the COVID-19 pandemic, the focus on disinformation campaigns has shifted from foreign

[13] Jack Snyder, "Backlash against naming and shaming: The politics of status and emotion", *The British Journal of Politics and International Relations* 22, no 4 (2020): 644-653.

threat actors to domestic ones.[14] On balance, disinformation campaigns originating from domestic actors are more enduring and damaging than those originating from foreign actors.[15] But, when domestic actors are involved, even if co-opted by foreign actors, the factors that bear on the decision to identify domestic actors differ in important ways from those that apply to foreign actors.

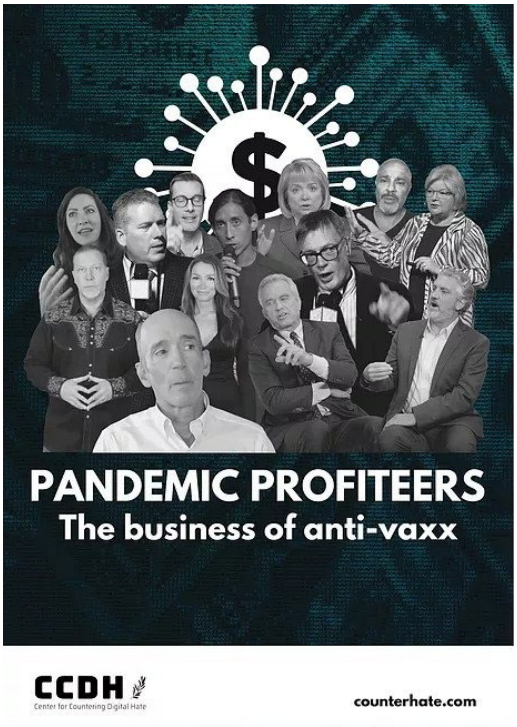


Figure 8 Pandemic Profiteers, Center for Countering Digital Hate.

Both private and public entities have legal obligations to protect information they collect about information consumers. This obligation depends on the type of information collected and how it was collected. Private entities generally have no legal obligation to protect information about information consumers with whom they have no fiduciary relationship, provided the information was gathered through licit means. Therefore, these private entities are free to publish this information when it suits their purpose. For example, in a report published in March 2021, the Center for Countering Digital Hate named twelve information consumers responsible for 73% of anti-COVID-19 vaccine content online.[16]

[14] Larry Luxner, "Ahead of the 2020 US elections, the disinformation threat is more domestic than foreign", Atlantic Council, last modified September 23, 2020, https://www.atlanticcouncil.org/blogs/new-atlanticist/ahead-of-the-2020-us-elections-the-disinformation-threat-is-more-domestic-than-foreign (accessed August 24, 2021).
[15] Richard Stengel, "Domestic Disinformation Is a Greater Menace Than Foreign Disinformation", *Time*, last modified June 26, 2020, https://time.com/5860215/domestic-disinformation-growing-menace-america (accessed August 24, 2021).
[16] Center for Countering Digital Hate, "The Disinformation Dozen: Why Platforms Must Act on Twelve Leading Online Anti-Vaxxers", PDF file, March 24, 2021, https://252f2edd-1c8b-49f5-9bb2-cb57bb47e4ba.filesusr.com/ugd/f4d9b9_b7cedc0553604720b7137f8663366ee5.pdf. (accessed August 24, 2021).

The report included the name, photo, and screenshots of posts on Facebook and Twitter of each individual. The New York Times later published a more detailed article on the person listed in the report as the greatest offender.[17]

Whether the identification of domestic threat actors by private entities helps to stem the spread of disinformation is unclear. While naming and shaming may be superficially appealing as a method of exacting a price from those who peddle in disinformation, such naming and shaming might make matters worse by provoking a fierce backlash from those sympathetic to the views of the information consumers identified. In the end, this further entrenches both sides in their respective ideological positions.[18] Since polarization is a primary reason for the success of disinformation campaigns, attempting to counter these campaigns with methods which may generate even more polarization seems questionable. Also, revealing the identities of domestic threat actors and other personally identifiable information can render these information consumers vulnerable to harassment or more egregious forms of retaliation.[19] We do not condone harassment and vigilantism as means of responding to domestic threat actors.

Absent suspicion of criminal activity or a legitimate government purpose pursuant to clear legal authority, government monitoring of the opinions and activities of U.S. persons is problematic. Although federal and state law penalize false statements in judicial proceedings and official documents and information consumers may be liable for words

[17] Sheera Frankel, "The Most Influential Spreader of Coronavirus Misinformation Online", *N.Y. Times*, last modified July 24, 2021, https://www.nytimes.com/2021/07/24/technology/joseph-mercola-coronavirus-misinformation-online.html (accessed July 30, 2021).
[18] Ibid.
[19] Kim Zetter, "Cyberbullying Suicide Stokes the Internet Rage Machine", *Wired*, last modified November 21, 2007, https://www.wired.com/2007/11/cyberbullying-suicide-stokes-the-internet-fury-machine (accessed August 24, 2021).

that incite violence, defraud, or defame, disinformation, much less misinformation, is not illegal in most cases.

2.1.3. Uncovering Imposters

An alternative to naming specific domestic threat actors is to alert information consumers to those threat actors who assume fake persona or claim to have credentials which they do not have. The status of the threat actor plays an important role in influencing the perceived trustworthiness of that threat actor.[20] Someone who claims to be an epidemiologist will likely be viewed as more reliable on the topic of coronaviruses than a person who claims to be a bus driver.

Information consumers who assume fake personas or claim fake credentials to deceive others for illegitimate purposes have no moral or legal standing for protection for their fraud, which is a form of disinformation. Revealing that credentials are fake can be accomplished without identifying the threat actor by name or by other information which can be tied to a specific person.

Figure 9 Fake identities, (obtained from Wonder How To).

[20] Edward L. Glaeser, "Measuring Trust", PDF file, *The Quarterly Journal of Economics*, (August 2000): 811-846. https://scholar.harvard.edu/files/laibson/files/measuring_trust.pdf (accessed August 24, 2021).

Threat actors who spread disinformation often seem to act with impunity, facing few negative consequences for the harm that they cause. In some circles, their disinformation may enhance their status and generate lucrative opportunities for them due to the attention that they draw. Legal actions against threat actors are available, but limited in number, and invariably costly and time-consuming. In a deeply divided society, shunning and ostracism no longer have the practical import they may have had in earlier generations. The lack of perceived consequences encourages threat actors to continue their activities. We acknowledge the challenges in today's environment and believe that punitive measures are less effective than measures which provide information about the threat actors upon which information consumers can make better informed decisions.

❖ Recommendation 2: Uncover Imposters

2.2. Social Media Algorithms & Fact-Checking

Challenging disinformation by fact-checking the information and making the results of this fact-checking publicly available is an important method of combatting disinformation campaigns. Disinformation which spreads unchecked through the information ecosystem will ultimately end up in the content feeds of social media platforms. Algorithms determine much of the content which ends up in these feeds. By allowing information consumers more control over these algorithms and the sources which these algorithms utilize, information consumers may have a greater ability to exclude known sources of disinformation.

Figure 10 Misinformation, (obtained from Agility PR Solutions).

2.2.1. Social Media Algorithms

Social media companies use algorithms to sort, index, prioritize, and sometimes suppress the content generated by users of their platforms. These companies can manipulate these algorithms to enhance the profitability of their platforms by attempting to keep users engaged on their platforms as much as possible through attention-grabbing features such as likes, comments, streaks, and recommended posts and people to follow. User activities are recorded, quantified, and used to tailor the user's experience while on the social media platform. On the one hand, algorithms simply give

the user more of what the user's activity on the platform seems to indicate the user wants. On the other hand, these same algorithms potentially create polarizing echo chambers by excluding contrary views.

In April 2021, the United States House of Representatives Subcommittee on Privacy, Technology, and the Law, a subcommittee of the Committee on the Judiciary, met to discuss the impact of algorithms.[21] Several panelists warned of the dangers of algorithms and broadly echoed the sentiments of others who claim that algorithms affect how we see the world[22], erode our ability to freely make choices[23], inherently create polarization[24], and create dangerous feedback loops[25]. The panelists offered solutions ranging from robust federal oversight to a new digital infrastructure which avoids the many problems created by social media platforms in their current forms.

Algorithms influence decision-making in many sectors of society from determining insurance rates to assessing the volatility of stock prices on Wall Street. Algorithms require data to function, which comes in the form of user input on social media platforms.

---

[21] *Algorithms and amplification: How Social Media Platforms' Design Choices Shape Our Discourse and Our Minds,* 117th Congress, Subcommittee on Privacy, Technology, and the Law, Committee on the Judiciary, U.S. Senate, April 27, 2021, https://www.judiciary.senate.gov/meetings/algorithms-and-amplification-how-social-media-platforms-design-choices-shape-our-discourse-and-our-minds (accessed August 24, 2021).

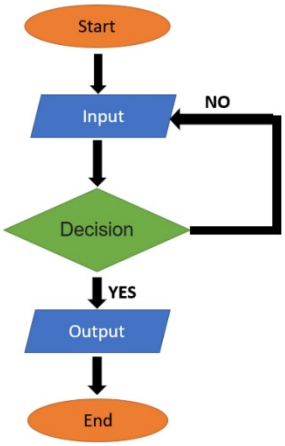
[22] Joanna Stern, "Social-Media Algorithms Rule How We See the World. Good Luck Trying to Stop Them, *Wall Street Journal*, last modified January 17, 2021, https://www.wsj.com/articles/social-media-algorithms-rule-how-we-see-the-world-good-luck-trying-to-stop-them-11610884800 (accessed August 2, 2021).

[23] Lewis Mitchell and James Bagrow, "Do social media algorithms erode our ability to makes decision freely? The jury is out", last modified October 11, 2020, https://theconversation.com/do-social-media-algorithms-erode-our-ability-to-make-decisions-freely-the-jury-is-out-140729 (accessed August 3, 2021).

[24] Charles Johnston, "How Social Media Platforms Inherently Create Polarization", *Psychology Today*, last modified November 29, 2020, https://www.psychologytoday.com/us/blog/cultural-psychiatry/202011/how-social-media-algorithms-inherently-create-polarization (accessed August 3, 2021).

[25] Ben Dickson, "What makes AI algorithms dangerous?", *TechTalks*, last modified June 10, 2020, https://bdtechtalks.com/2020/06/10/ai-weapons-of-math-destruction (accessed August 3, 2021).

Input into algorithms helps to determine the output of these algorithms. Social media platforms ultimately determine what content the algorithms feed into the content streams of users. By providing control of the data that the algorithm uses, platforms may help to ensure users have control over their individual experiences on the platform. Regular reminders to review the information that the algorithms use and make adjustments, if necessary, can help keep users engaged in the process of determining the content that shows up in their content feeds.



❖ Recommendation 3: Grant Users More Control over Content Feeds

2.2.2. Fact-checking

In general, publishers of information have an ethical, legal, and fiduciary responsibility to ensure that they publish information that is accurate. When mistakes are uncovered, publishers can retract articles and posts, send amended versions, and publicly announce the course of action taken to remedy the situation. Notable exceptions are media sites whose content is for satirical purposes or entertainment shows that generate fake news for comic effect. Social media companies are not publishers in the traditional sense of the word. These companies provide platforms where users of these platforms can add content. The user-driven content on these

platforms does not imply that the social media companies endorse the content. While these companies usually grant users the ability to moderate content on the pages and groups formed by the users themselves, content is not first subject to pre-approval by the social media companies prior to posting.

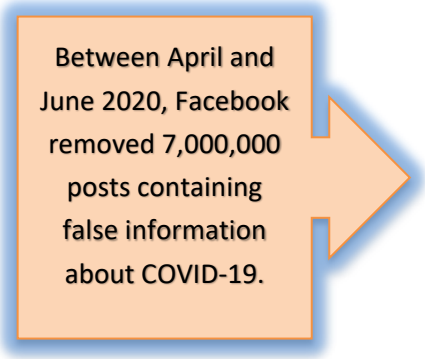


Figure 12 Reuters.

Social media companies arguably play an outsized role in providing access to news and providing a means to share news and other types of information. The power to determine what content is permissible on a social media platform and what is not is a power that social media companies should exercise with great discretion. These determinations should be made based on impartial and easily understood guidelines. Alerting users to the accuracy of information appearing in content feeds plays a useful role in encouraging users to make informed decisions.

Social media platforms, third-party entities, and users themselves can play a role in fact-checking information. If information is determined to be false, the information can be labeled as such, deleted, or supplemented by facts. To minimize the politicization of the process of removing or correcting disinformation, clear distinctions should be drawn between fact-checking, content moderation, and what constitutes improper censorship.

The goal of fact-checking is not to promote or suppress points of view, but to make sure that the information presented can be supported by verifiable information. We urge all fact-checkers to adhere to the principles identified below:

Figure 13 International Fact Checking Code of Principles, IFCN.

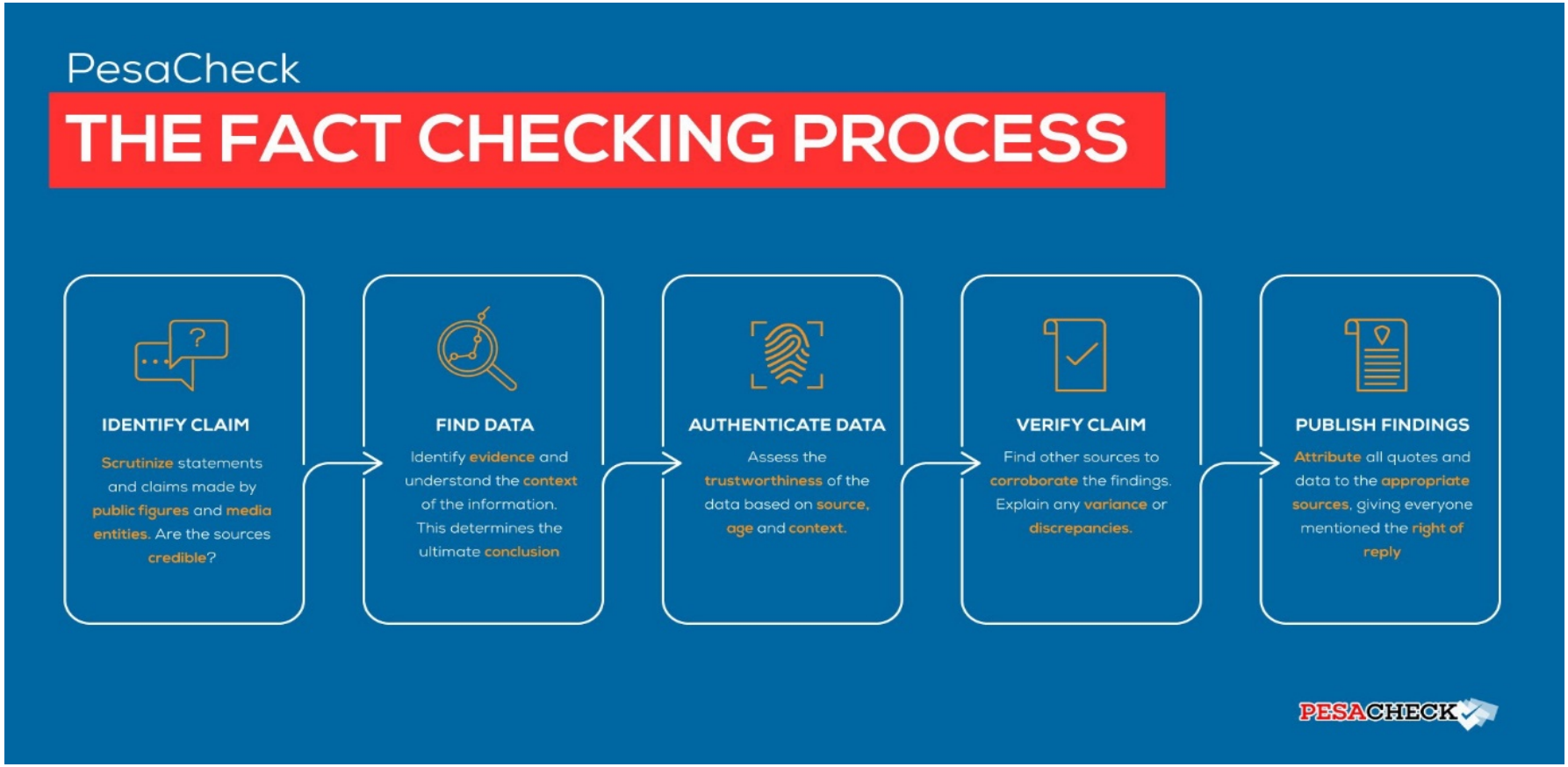
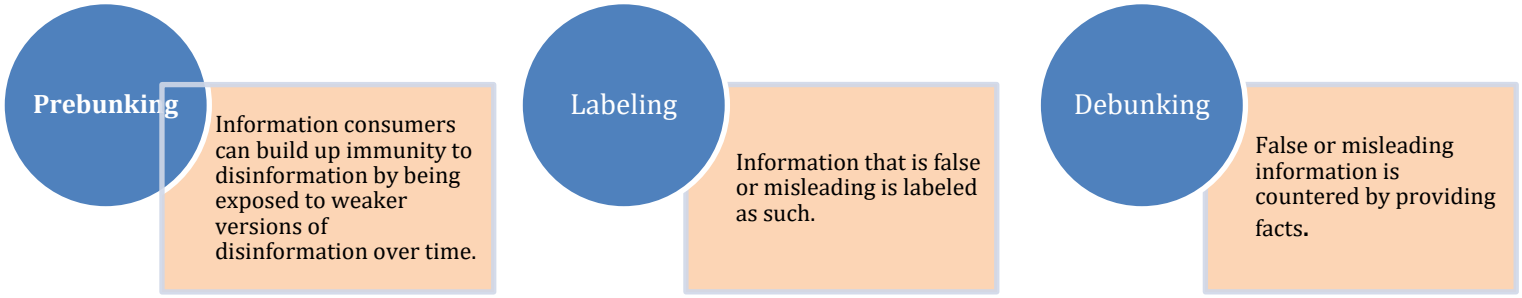


Figure 14 The Fact Checking Process, (obtained from PesaCheck).

There are different methods of notifying readers that information has been fact-checked, but found to be unverifiable or false.



Researchers have found that the timing of fact-checks provided to readers makes a difference.[26] In one study, participants were exposed to 18 true headlines and 18 false headlines with true and false tags before (prebunking), during (labeling), and after (debunking) reading the headlines. The participants rated the accuracy of each headline. A week later, the participants were asked again to rate the accuracy of each headline. In comparison to prebunking and labeling, debunking had the greatest impact on their ability to discern the truthfulness of the headlines.

More research needs to be conducted into the efficacy of one method versus another, particularly on a longer time scale. However, having accurate information does not necessarily lead to a change in belief.[27] Some researchers maintain that the effect is

[26] Nadia M. Brashier et al., "Timing matters when correcting fake news", *PNAS* 118, no 5 (2021): e2020043118. https://doi.org/10.1073/pnas.2020043118 (accessed August 24, 2021).
[27] Oscar Barrera et al., "Facts, alternative facts, and fact checking in times of post-truth politics", *Journal of Public Economics* 182, (2020). https://doi.org/10.1016/j.jpubeco.2019.104123 (accessed August 24, 2021).

weak.[28] Overall, the impact of fact-checking alone is limited.[29] The ability to spread disinformation far outpaces the ability to fact-check. Therefore, fact checkers need to prioritize where they can best employ their resources. We conclude that challenging misinformation and disinformation online is preferable to allowing both to flow unchecked.

❖ Recommendation 4: Increase Fact-Checking Efforts

3. The Demand for Disinformation

Disinformation campaigns wreak havoc because information consumers consume the disinformation, share it with others, and act upon it. For a variety of psychological and social reasons, information consumers are tempted to believe disinformation without weighing whether the disinformation is supported by evidence or sound reasoning.

> *A cognitive bias is a systemic error in thinking that occurs when people are processing and interpreting information in the world around them and affects the decisions and judgments that they make.*
>
> *Kendra Cherry*

Understanding the influence of cognitive biases and why information consumers make different choices about what to believe and share

[28] Nathan Walter, et. al., "Fact-checking: A Meta-Analysis of What Works and for Whom", *Political Communication* 37, no 3 (2020): 350-375, DOI: 10.1080/10584609.2019.1668894 (accessed August 24, 2021).
[29] Andrew Tompkins, "Is fact-checking effective? A critical review of what works – and what doesn't", *DW AKADEMIE*, last modified December 10, 2020, https://www.dw.com/en/is-fact-checking-effective-a-critical-review-of-what-works-and-what-doesnt/a-55248257 (accessed August 2, 2021).

with others can help make them more resilient to threat actors who exploit the way information consumers think and interact with one another.

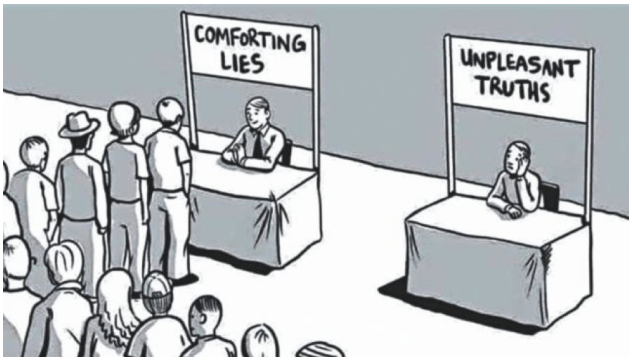


Figure 15 Comforting Lies vs. Unpleasant Truths (obtained from News Literacy Matters).

Threat actors are not alone in attempting to use the lessons of cognitive science and group psychology to their advantage. Governments, religious authorities, and retail advertisers have long taken advantage of psychological and social characteristics of information consumers to induce conformity, sell products, discourage harmful health habits, and otherwise steer thinking and behavior in a desired direction. Information consumers use reverse psychology or play on known weaknesses of others to achieve certain aims. Information consumers are not always aware of these efforts to manipulate them.

The information ecosystem has a profound impact on how we view the world. An information ecosystem contaminated by disinformation can influence people to think, feel, and behave in ways not informed by evidence, but by the duplicity of determined threat actors.

> An environment is, after all, a complex message system which imposes on human beings certain ways of thinking, feeling, and behaving. It structures what we can see and say and, therefore, do. It assigns roles to us and insists on our playing them. It specifies what we are permitted to do and what we are not.
>
> Neil Postman

In this section, we conclude that information consumers can become more resilient to disinformation online by understanding how their brains process information, improving

their ability to think critically, and becoming aware of how their online activity can facilitate the spread of disinformation.

3.1. Cognitive Bias

The mechanisms by which human brains process information make information consumers vulnerable to disinformation. For the human brain to process the enormous amounts of sensory input it receives; the brain has evolved mental shortcuts or "cognitive biases". However, these cognitive biases can undermine the ability to identify relevant facts, weigh the relevance of these facts, and form coherent courses of action based on these facts. Confirmation bias, belief bias, and the bandwagon effect are cognitive biases that are particularly relevant to understanding disinformation campaigns.

Confirmation bias

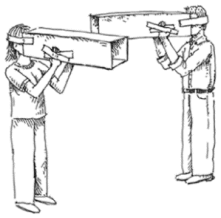


Figure 16 Tunnel Vision (obtained from Aftercare.com).

Information consumers who rely primarily on sources of information that conform to their ideological preferences are particularly vulnerable to disinformation. The design of online search engines, social media platforms, and smartphone applications, plus the availability of cable news programming catering to specific audiences, make it easy to screen information consumers from viewpoints that conflict with their ideological preferences.[30]

[30] Silvia Knobloch-Westwick and Steven B. Kleinman, "Preelection Selection Exposure: Confirmation Bias Versus Informational Utility", *Communication Research* 39, no. 2 (2012): 170-193. https://doi.org/10.1177%2F0093650211400597 (accessed August 24, 2021).

Belief bias

Figure 17 See hear speak no evil cartoon (obtained from VectorStock).

Information consumers tend to favor data and arguments which support their predetermined conclusions and view more harshly data and arguments which do not. While researching the spread of disinformation prior to the 2016 presidential election, researchers determined that whether a news item was accepted as true or rejected as false by information consumers strongly depended on how much it conformed to their belief system.[31]

Bandwagon effect

Figure 18 Blind Leading the Blind (obtained from Conversion Uplift).

Information consumers tend to adopt the beliefs that other people in their political and social networks have. Research has shown that exposure to pre-election polls increases the likelihood that voters will side with majority opinions.[32]

Cognitive biases facilitate analytical errors, lapses in judgment, faulty conclusions, hasty generalizations, and other defects in sound reasoning. Since cognitive biases are features of the human brain, learning how to minimize their impact by developing critical things skills is essential to building resilience to disinformation.[33]

---

[31] Giovanni Luca Ciampaglia and Filippo Menczer, "Misinformation and bias infect social media, both intentionally and accidentally", *The Conversation*, last modified June 20, 2018, https://theconversation.com/misinformation-and-biases-infect-ocial-media-both-intentionally-and-accidentally-97148 (accessed August 24, 2021).
[32] Mike Farjam, "The Bandwagon Effect in an Online Voting Experiment With Real World Political Organizations", PDF file, *International Journal of Public Opinion Research vol 33 (2)* (Summer 2021): 412-421. https://doi.org/10.1093/ijpor/edaa008 (accessed August 24, 2021).
[33] Paul Machete and Marita Turpin, "The Use of Critical Thinking to Identify Fake News: A Systematic Literature Review", PDF file, *IFIP International Federation for Information Processing*, (2020): 235-246. https://link.springer.com/content/pdf/10.1007%2F978-3-030-45002-1_20.pdf (accessed August 24, 2021).

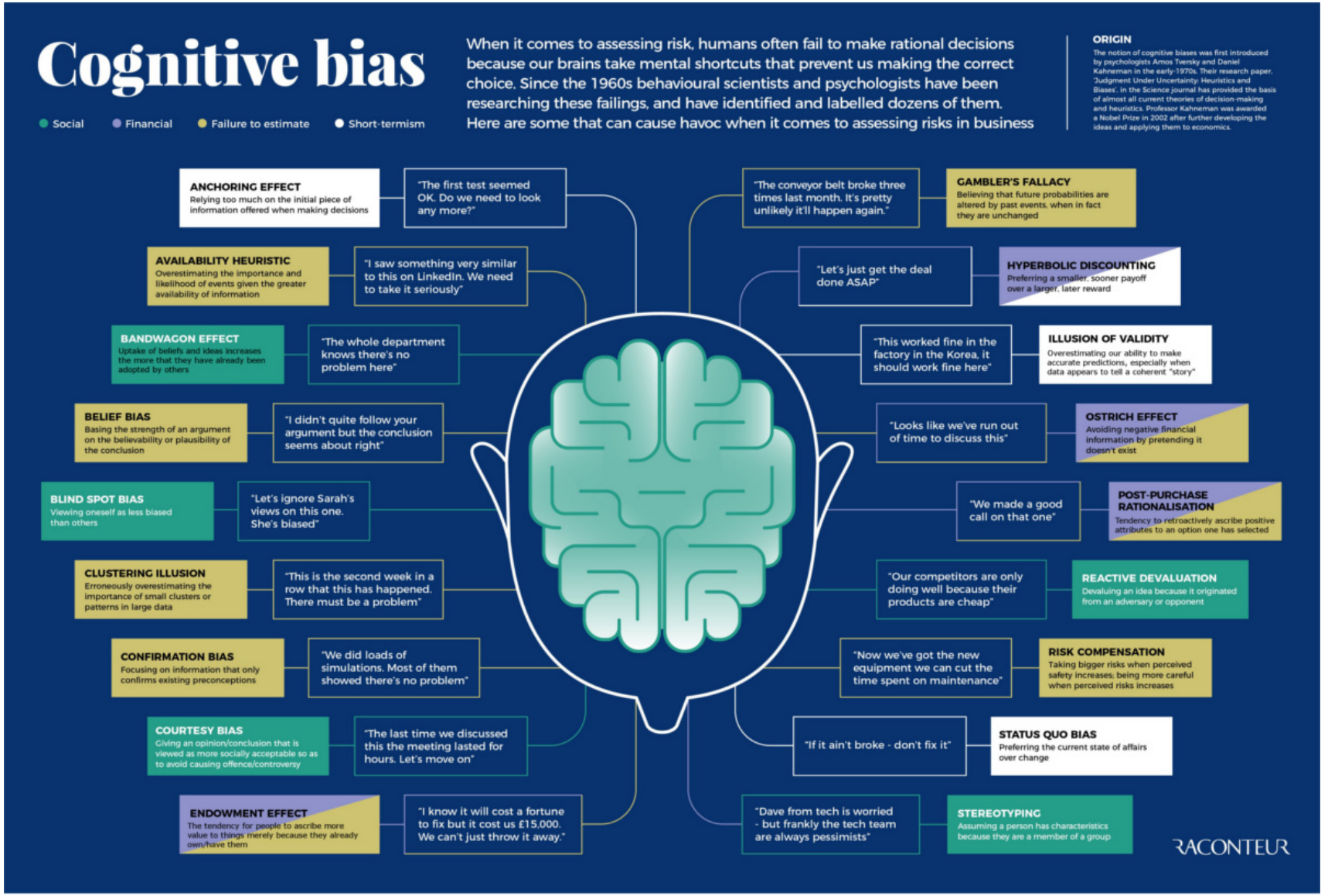


Figure 19 Cognitive Bias, [Raconteur](#).

3.2. Emotional reactions to disinformation

Strong emotions influence the ability of information consumers to process information. Information that elicits strong emotions can overwhelm an individual's ability to think rationally and make prudent decisions. Strong emotions fuel "outrage culture" and lend their force to the incessant volleys of mutual recriminations which are commonplace on social media platforms. Threat actors start disinformation campaigns not to inform, but to agitate, provoke, incite, and inflame. By doing so, they render their audiences more receptive to the disinformation they wish to promote.

Figure 20 Angry Comments on Social Media (obtained from ISM Works).

The belief in the truthfulness of fake news is largely dependent on whether the information consumer relies primarily on reason or emotion to assess the information.[34] The more the information consumer relies on emotion, the more likely the information consumer is to believe in the fake news. Additionally, information consumers are more likely to share information if that information provokes emotional reactions.[35] Researchers determined that the ability to attract an information consumer's attention is key to whether the information consumer will later share this information. Information that does not elicit strong emotions is less

[34] Cameron Martel, Gordon Pennycock, and David G. Rand, "Reliance on emotion promotes belief in fake news", *Cognitive Research: Principles and Implications*. 5:47 (2020). https://doi.org/10.1186/s41235-020-00252-3 (accessed August 24, 2021).

[35] William J. Brady et al., "Emotion shapes the diffusion of moralized content in social networks", *PNAS* 114, no. 28 (2017): 7313-7318. Doi: 10.1073/pnas.1618923114 (accessed August 24, 2021).

likely to attract the attention of information consumers and therefore less likely to be shared.

The circumvention of thoughtful deliberation before acting, rather than the presence of strong emotions, is the key issue. Information consumers often make poor decisions in the heat of the moment, and when blinded by outrage or the desire for revenge. When information consumers encounter information online which elicits such reactions, these feelings should give the individual pause.

3.3. The Psychology Behind Sharing Information Online

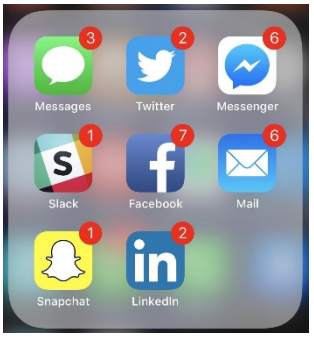
Information consumers who share disinformation help sustain disinformation campaigns and contribute to the rapid spread of disinformation through the information ecosystem. Research suggests that relatively few people intentionally share disinformation on social media.[36] Three main factors that drive the decision to share disinformation are consistency, consensus, and authority.[37] First, the disinformation is consistent with the beliefs that the information consumers already possess. Second, the information consumer believes that most people in his or her social group believe the disinformation. Third, the information consumer believes that the disinformation derives from an authoritative source.

[36] Tom Buchanon, "Why Do People Share Disinformation on Social Media?", PDF file, *Policy Brief*, September 2020, https://crestresearch.ac.uk/download/3040/20-017-03-disinformation-on-social-media.pdf (accessed August 24, 2021).

[37] Tom Buchanon, "Why Do People Spread Disinformation Online: The effects of message and viewer characteristics on self-reported likelihood of sharing social media disinformation", *PLoS ONE* 15, no 10 (2020): e0239666. https://doi.org/10.1371/journal.pone.0239666 (accessed August 24, 2021).

Figure 21 Eric Allie/Cagle Cartoons.

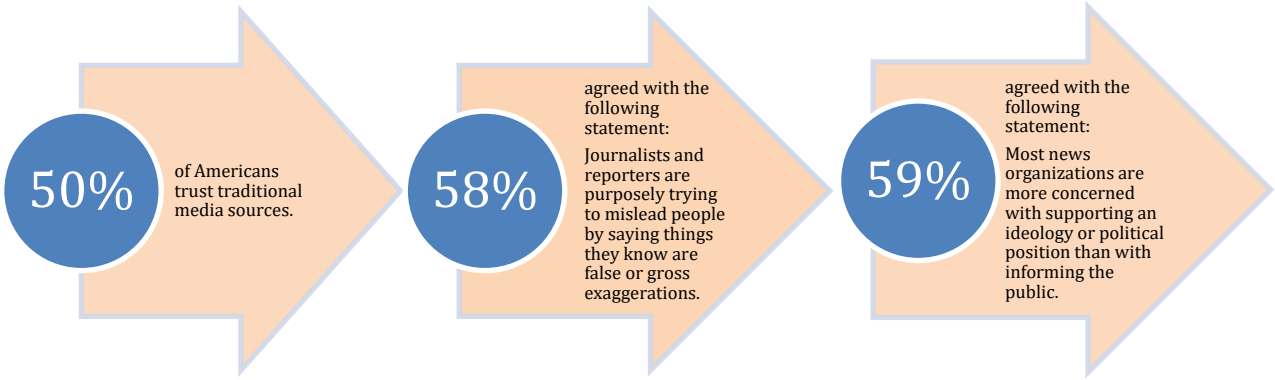
Other factors also play a role in the decision to share disinformation. Information consumers may share disinformation on social media platforms to demonstrate allegiance to a particular idea (similar to virtue signaling), to share novel information, or to draw attention to themselves to increase 'likes', 'views', or 'followers'. Political partisanship motivates information consumers to share disinformation when this disinformation can be used to counter the arguments of political opponents.[38]



Social media platforms have features that incentivize information consumers to share information. These features often have addictive qualities that make it difficult for information consumers to resist the temptation to share information, including disinformation, or to disengage from the platform entirely. Before sharing information online, we encourage information consumers to critically examine the information they wish to share and the motives they have for sharing it. Disinformation at rest is far less effective than disinformation in motion.

[38] Mathias Osmundsen et al, "Partisan Polarization is the Primary Psychological Motivation behind Political Fake News Sharing on Twitter," *American Political Science Review* 115, no 3, (2021): 1-17.

❖ **Recommendation 5: Teach Information Consumers about the cognitive and emotional impact of social media and online information**



Edelman Trust Barometer 2021.

4. Media Literacy and Critical Thinking Skills

Media literacy is the application of critical thinking skills to forms of information, whether print newspapers and magazines, along with their online counterparts, broadcasts on television and radio, online video or podcasts, and on social media platforms or messaging apps. Media literacy efforts typically involve training information consumers how to differentiate fact from opinion, how to assess sources of information, and how information can be manipulated to deceive the audience. Research suggests

that media literacy efforts can assist information consumers to distinguish real news from fake news.[39]

Other approaches include helping information consumers understand the cognitive and mental health impact of using social media and other digital sources of information and helping them negotiate the many hazards found online.[40] One researcher maintains that media literacy efforts should encompass not only traditional forms of media, but also the information that passes through their networks of online connections.[41] Using this approach, information consumers can more clearly appreciate their own roles in the dissemination of disinformation. An article, video, podcast, or photograph shared, a retweet, a like, a comment, a post to a listserv, all may have more impact than an individual might foresee.

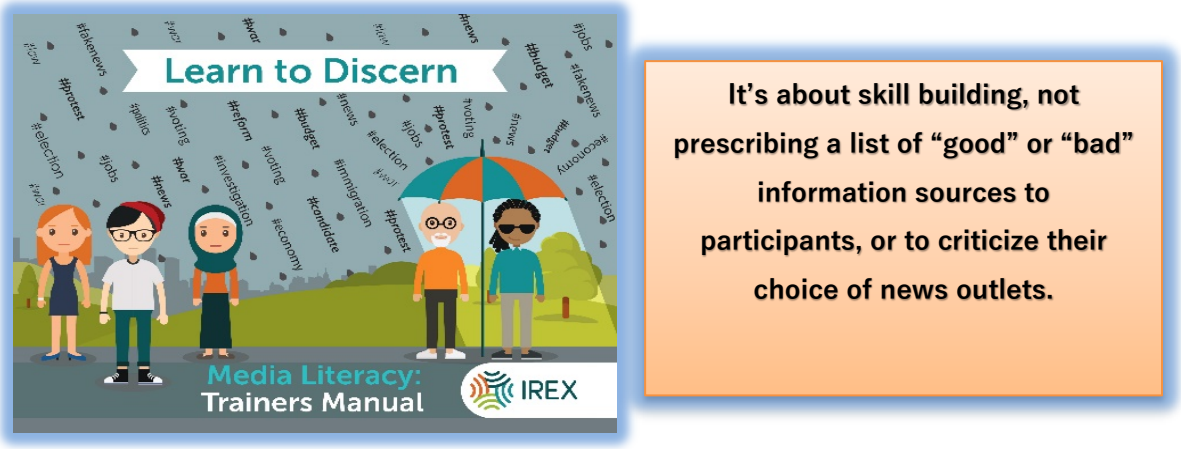


Figure 22 IREX Media Literacy Training Manual.

[39] Andrew Guess et al., "A digital media intervention increases discernment between mainstream and false news in the United States and India", PDF file, PNAS Vol. 117, no.27 (2020): 15536-15545. https://www.pnas.org/content/pnas/117/27/15536.full.pdf (accessed August 24, 2021).
[40] Amy Callahan, "Media Literacy Isn't Coming to Save Us (But We Can Make It Better)", *EdW*, last modified January 23, 2019, https://www.edweek.org/teaching-learning/opinion-media-literacy-isnt-coming-to-save-us-but-we-can-make-it-better/2019/01 (accessed August 24, 2021).
[41] Dana Boyd, "You Think You Want Media Literacy…Do You?," *Points: Data & Society*, last modified March 9, 2018, https://points.datasociety.net/you-think-you-want-media-literacy-do-you-7cad6af18ec2 (accessed August 24, 2021).

Educational institutions and non-governmental organizations provide resources on media literacy for use in formal and information educational settings, and self-paced learning. In 2019, the Digital Citizenship and Media Literacy Act was introduced in the U.S. Senate to assist with nationwide efforts, including the study of best practices in media literacy programs.[42]

Research indicates that many information consumers are not able to articulate the criteria by which they evaluate the trustworthiness of information and therefore are unable consistently to distinguish real news from fake news.[43] The development of critical thinking skills can help information consumers recognize faulty reasoning, weakly-

Figure 23 Critical Thinking Diagram (obtained from Tycoonstory).

[42] A Bill to promote digital citizenship and media literacy, S. 2240, 116th Congress, 1st Session. (2019).
[43] Blanca Puig, Paloma Blanco-Anaya, and Jorge J. Pérez-Maciera, "'Fake News' or Real Science? Critical Thinking to Assess Information on COVID-19", *Frontiers in Education*, last modified May 3, 2021, https://www.frontiersin.org/articles/10.3389/feduc.2021.646909/full (accessed August 24, 2021).

supported arguments, and disinformation. Teaching critical thinking skills as a separate course, as well as incorporating the application of critical thinking in all courses, leads to better outcomes than not providing a course focused exclusively on the teaching of critical thinking skills.[44] We encourage the development of critical thinking skills in both formal and informal educational settings.

❖ Recommendation 6: Teach Media Literacy and Critical Thinking Skills

We also suggest information consumers to utilize the SMART mnemonic and practices. Based on Aesop's The Tortoise and the Hare, the SMART graphic on the next page illustrates the need to take time to reflect on information to verify its truthfulness before dissemination to others. Proceeding slowly, but prudently, before acting on information is preferable to proceeding rapidly, but carelessly. The tortoise bypasses the hurdles that impede the search for the facts. Meanwhile, the hare falls headlong into the traps set by those who promote disinformation.

[44] Shane Horn and Koen Veermans, "Critical Thinking efficacy and transfer skills defend against 'fake news' at an international school in Finland," *Journal of Research in International Education*, 18, no. 1 (2019): 23-41. https://doi.org/10.1177%2F1475240919830003 (accessed August 24, 2021).

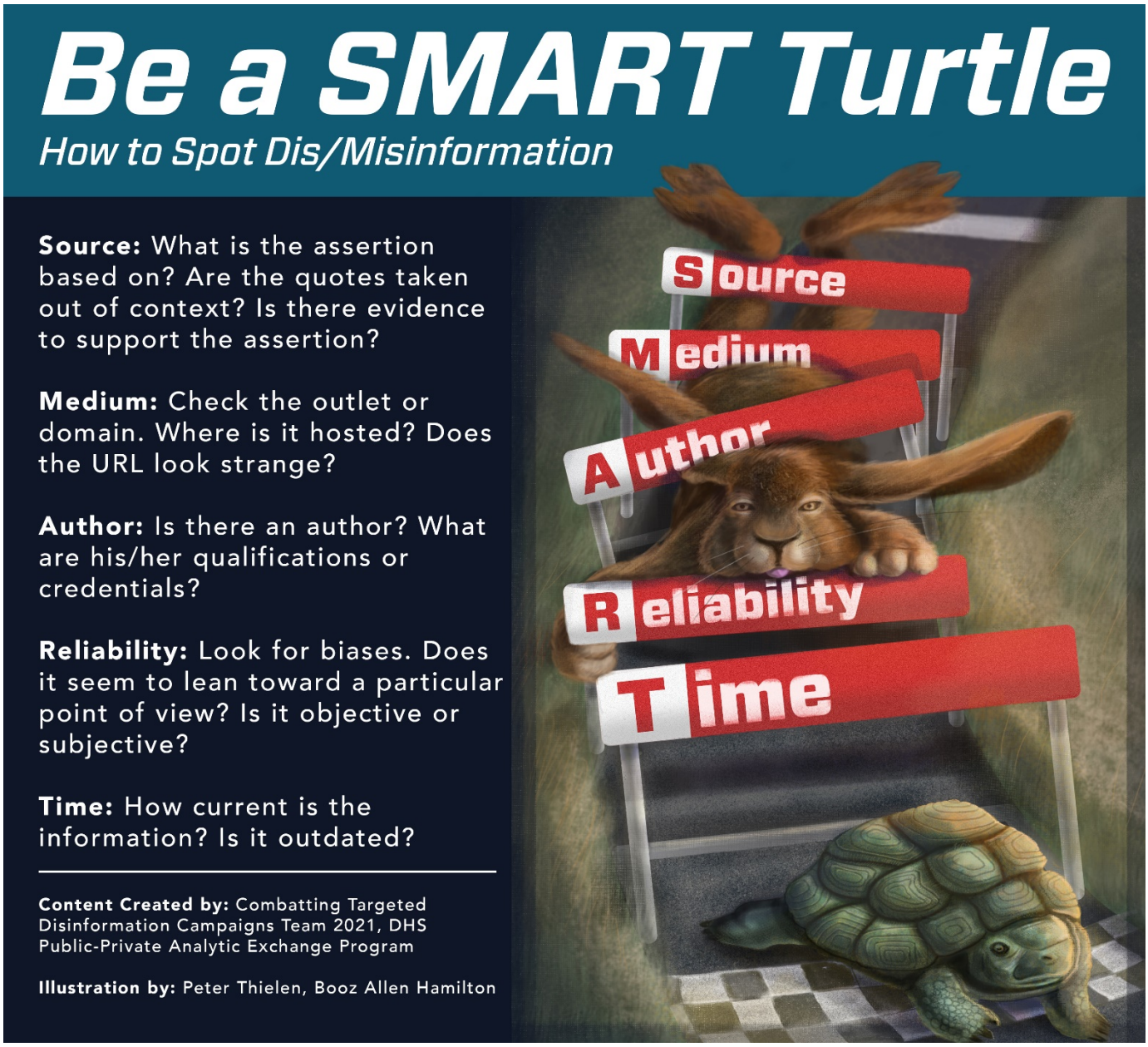


Figure 24 SMART Graphic | Source: Combatting Targeted Disinformation Campaigns Team 2021

Conclusion

It is extremely improbable that disinformation campaigns will disappear in the foreseeable future. Despite the increasing effectiveness of countermeasures, threat actors will always find avenues to spread their disinformation and will adopt new tactics as new forms of technology emerge. Compared to many legitimate forms of persuasion and influence, disinformation campaigns are inexpensive and frequently have few downsides for the threat actor. Threat actors take ready advantage of software and communication platforms that they did not develop and benefit from political and social conditions that they did not create.

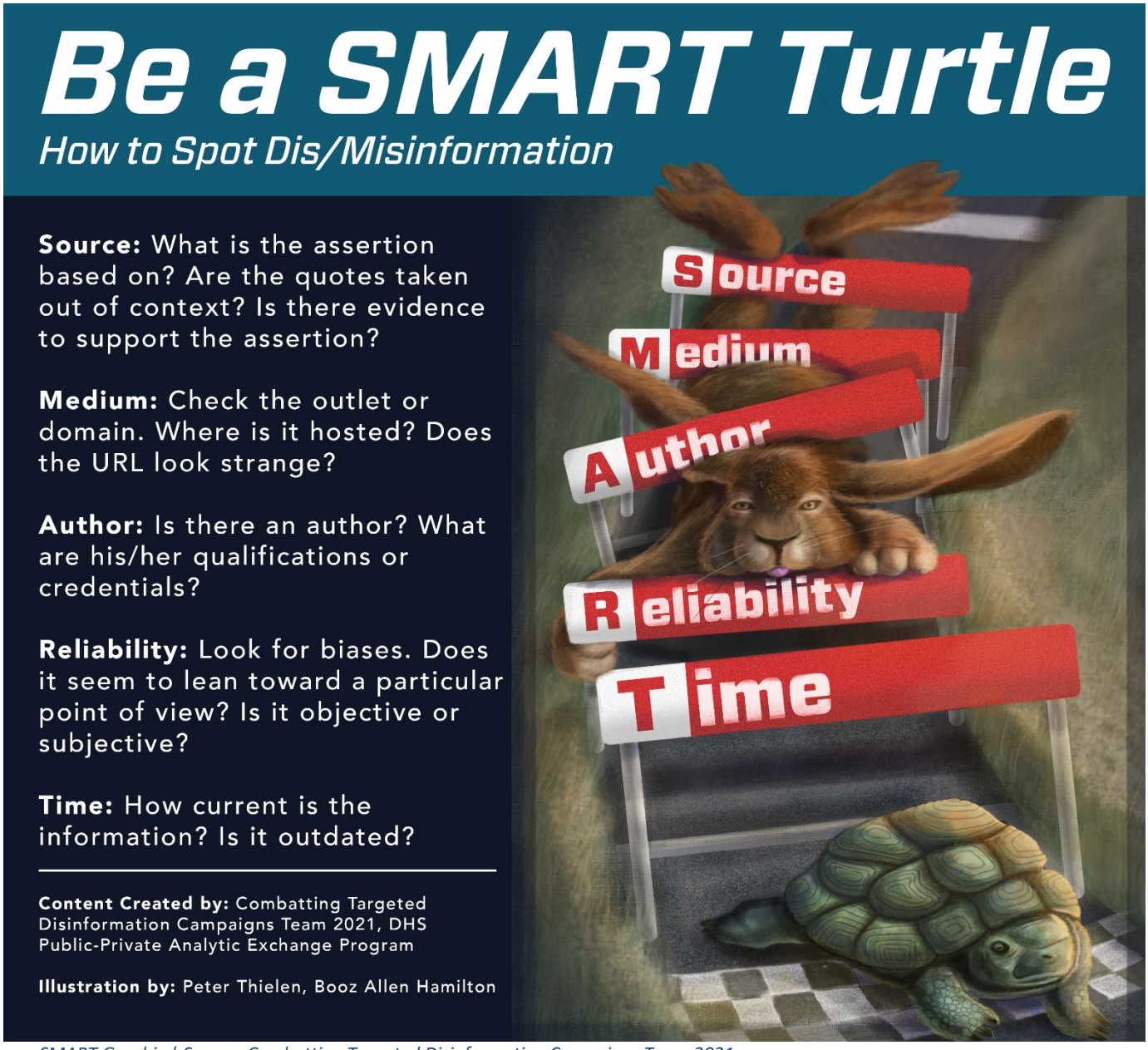
In this report, we concluded that disinformation campaigns are threats to national security because they undermine the well-being of our society. Though disinformation campaigns cannot be stopped fully, we believe that measures can be taken to impede these campaigns and reduce their impact. We believe that building the resilience of information consumers to disinformation will likely bear more fruit than focusing on technological solutions. To build such resilience, we believe that giving information consumers more tools with which they can verify the information they consume online, identify the threat actors behind disinformation campaigns, verify the claims of imposters hiding behind fake persona and credentials, and control their content feeds is essential.

The factors which make information consumers vulnerable to disinformation are rooted in human psychology, the divisions within our society, and the siloed nature of today's information ecosystem. As illustrated in the Disinformation Kill Chain, one result of

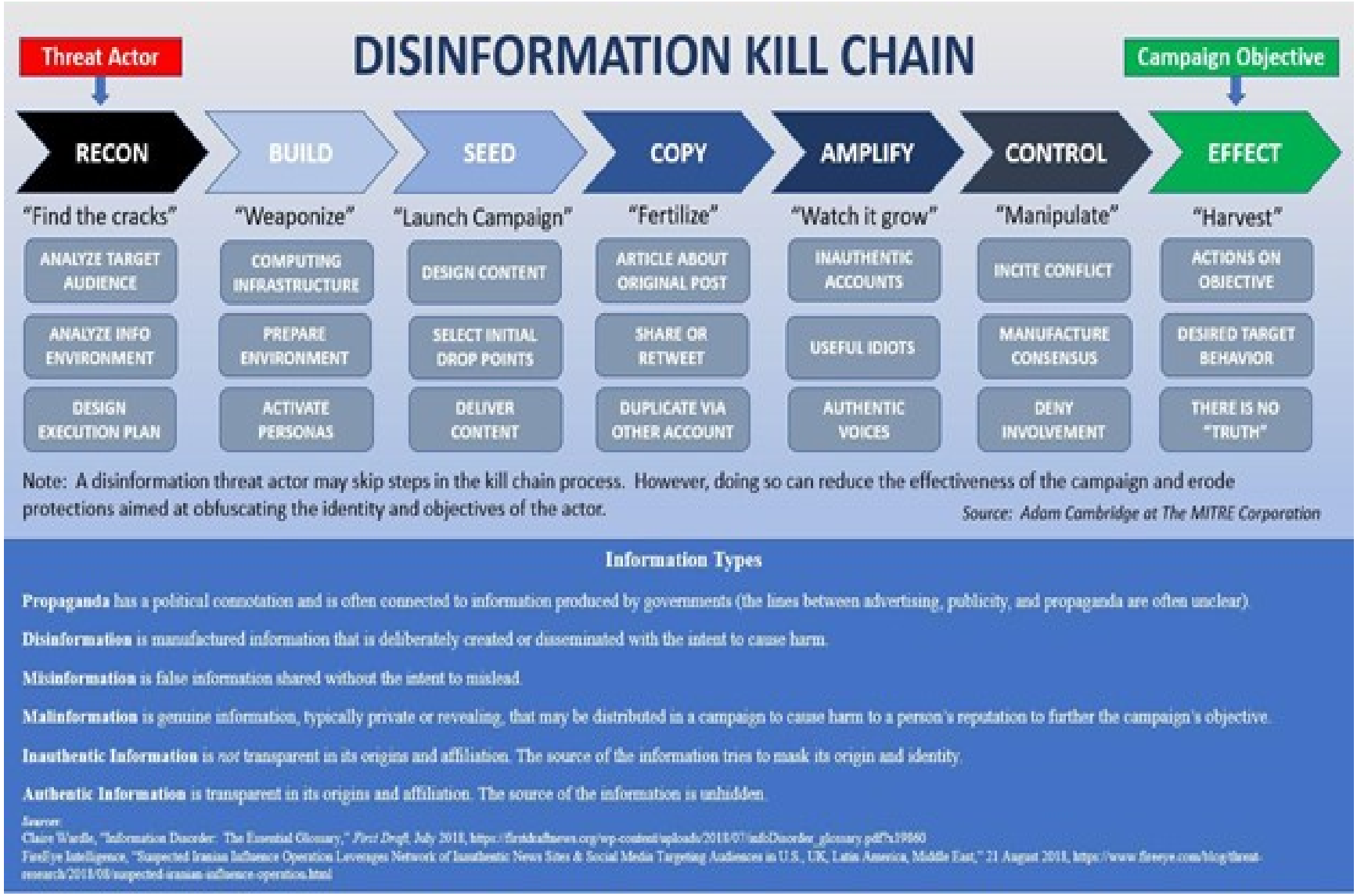
disinformation campaigns is that information consumers end up believing that truth no longer exists. Such a result demonstrates that supply and demand form a positive feedback loop. When information consumers lose the ability to distinguish fact from fiction, and become uninterested in doing so, demand for disinformation grows, which drives the supply of disinformation.

The prevalence of disinformation campaigns in our society is emblematic of the polarization of our society. Such polarization hampers a united response and even the ability to come to a common understanding of what a fact is, and what disinformation or misinformation are. Individual information consumers can build up their immunity to disinformation. But our society, as a whole, will not be able to build up its resilience until the larger problems that recent events have exposed are dealt with first.

Exhibits



SMART Graphic | Source: Combatting Targeted Disinformation Campaigns Team 2021.



Source: The MITRE Corporation.

DISCLAIMER STATEMENT: This document is provided for educational and informational purposes only. The views and opinions expressed in this document do not necessarily state or reflect those of the United States Government or the Public-Private Analytic Exchange Program, and they may not be used for advertising or product endorsement purposes. All judgments and assessments are solely based on unclassified sources and are the product of joint public and private sector efforts.