USER CORRECTION AS A TOOL IN THE BATTLE AGAINST SOCIAL MEDIA MISINFORMATION

Leticia Bode*

CITE AS: 4 GEO. L. TECH. REV. 367 (2020)

TABLE OF CONTENTS

I. INTRODUCTION

Misinformation is not a new problem. As long as information is valuable in helping people make decisions, there will be an incentive for third parties to manipulate that information in a way beneficial to their interests. The result is misinformation, which has taken many forms over the years.

Misinformation has a variety of definitions. In this Paper, misinformation is used to mean information that is "considered incorrect based on the best available evidence from relevant experts at the time."[1] This definition does not discriminate based on intent—information is incorrect,

* Provost's Distinguished Associate Professor in the Communication, Culture, and Technology program at Georgetown University. Dr. Bode's research focuses on political and health communication, particularly in terms of how technology affects whether people are exposed to information, how accurate it is, and what they do with that information. Ph.D. University of Wisconsin—Madison, 2012. The author would like to thank Emily Vraga for years of thinking through many of these ideas.

[1] Emily K. Vraga & Leticia Bode, *Defining Misinformation and Understanding Its Bounded Nature: Using Expertise and Evidence for Describing Misinformation*, 37 POL. COMM. 136, 138 (2020).

regardless of whether its intent is to mislead[2] (disinformation) or not (misinformation).[3]

Although misinformation itself is not new, the form it has taken in recent years has changed. Technological innovation, broadly speaking, and social media, more specifically, allows for the rapid dissemination of both information and misinformation, and there is at least some evidence that the latter spreads more easily and more quickly than the former.[4] Although the nature of the problem may be overstated—a recent study found that only 8.5% of people shared misinformation with their friends on social media[5]—*perceptions* that there is a misinformation problem on social media are widespread. A recent poll illustrates the issue, showing that 82% of Americans believe they will read misleading information on social media.[6] Most people (59%) think that it is hard to tell the difference between factual and misleading information, and despite years of trying to fix this problem, most people (55%) actually think this task will become more difficult leading up to the 2020 election than it was in the 2016 election.[7]

Thus, blame for the misinformation problem is placed squarely on social media. Critics have therefore suggested that *the solution* to the problem lies with social media. If misinformation is rampant on Facebook, Twitter, Google, and other social platforms, then surely these platforms should use their technology to repair our flawed information environment. Social media platforms themselves have been eager to turn to technology—namely in the form of automated content moderation—as a means of addressing the problem.

Solutions from technology platforms generally focus on the ability of artificial intelligence and machine learning to identify and remove problematic content. This focus tends to follow the "A, B, C" model of content moderation,

---

[2] *See generally* Claire Wardle, *Fake News. It's Complicated.*, FIRST DRAFT NEWS (Feb. 16, 2017), https://firstdraftnews.org/latest/fake-news-complicated/ [https://perma.cc/39B6-QB7F] (distinguishing between different types of misinformation).

[3] Samuel Spies, *Defining "Disinformation"*, MEDIAWELL, (Oct. 22, 2019), https://mediawell.ssrc.org/literature-reviews/defining-disinformation/versions/1-0/ [https://perma.cc/L7M9-6RE8].

[4] *See* Soroush Vosoughi et al., *The Spread of True and False News Online*, 359 SCI. 1146, 1146–48 (2018).

[5] Andrew Guess et al., *Less Than You Think: Prevalence and Predictors of Fake News Dissemination on Facebook*, 5 SCI. ADVANCES 1, 1 (2019); *see also* Hunt Allcott & Matthew Gentzkow, *Social Media and Fake News in the 2016 Election*, 31 J. ECON. PERSP., 211, 213 (2017) (showing that people recall only about one false news story from the 2016 election).

[6] Brett Neely, *NPR Poll: Majority of Americans Believe Trump Encourages Election Interference*, NPR (Jan. 21, 2020, 5:01 AM), https://www.npr.org/2020/01/21/797101409/npr-poll-majority-of-americans-believe-trump-encourages-election-interference [https://perma.cc/NK6Z-BQ2L].

[7] *Id.*

so-called for its focus on manipulative Actors, deceptive Behaviors, and harmful Content.[8] Different pieces of data—including information about the person sharing the post (the "actor"), the context of actions surrounding the share (the "behavior" of the poster), and the post itself (the "content")—are input into a machine learning algorithm which is taught to classify content as either true or false.[9] Content that is flagged as false—or likely false—is either removed from the platform entirely or demoted in such a way that fewer people actually have the opportunity to view it.[10]

The technology companies in question are, indeed, pursuing these types of solutions. Most platforms have at least some public policy about how they deal with misinformation, and most of them use some form of automation to flag, identify, and remove or de-amplify false or misleading content and actors.[11]

But here lies the problem. Quite simply, it is one of scale. Given the sheer volume of information transmitted through social media, technology for content moderation cannot possibly solve the problem alone. Nor, I will argue, might we want it to do so.

The remainder of this Paper highlights some problems with technology-driven content moderation and proposes an alternative approach to dealing with misinformation at scale on social media.

---

[8] *Online Imposters & Disinformation: Hearing Before the Subcomm. on Investigations & Oversight of the H. Comm. on Sci., Space, and Tech.*, 116th Cong. 2, 4, 6 (2019) (testimony of Camille Francois, Chief Innovation Officer, Graphika), https://science.house.gov/imo/media/doc/Francois%20Addendum%20to%20Testimony%20-%20ABC_Framework_2019_Sept_2019.pdf [https://perma.cc/49C3-JCK6].

[9] *Id.; see also* Limeng Cui et al., *dEFEND: A System for Explainable Fake News Detection*, 28 PROC. ACM INT'L CONF. ON INFO. AND KNOWLEDGE MGMT. 2961 (2019).

[10] Guy Rosen & Tessa Lyons*, Remove, Reduce, Inform: New Steps to Manage Problematic Content*, FACEBOOK (Apr. 10, 2019), https://about.fb.com/news/2019/04/remove-reduce-inform-new-steps/ [https://perma.cc/45Z4-3FVP].

[11] *See, e.g., id.*; Colin Crowell, *Our Approach to Bots and Misinformation*, TWITTER (June 14, 2017), https://blog.twitter.com/en_us/topics/company/2017/Our-Approach-Bots-Misinformation.html [https://perma.cc/6LPE-7HT8]. An important caveat to this has emerged in the last year, wherein at least one major player in this space—Facebook—has said it will not monitor the content of political ads by candidates running for office, on the premise that such an approach is "grounded in Facebook's fundamental belief in free expression, respect for the democratic process, and the belief that, especially in mature democracies with a free press, political speech is the most scrutinized speech there is. Just as critically, by limiting political speech we would leave people less informed about what their elected officials are saying and leave politicians less accountable for their words." *See Fact-Checking on Facebook: What Publishers Should Know,* FACEBOOK (2020), https://www.facebook.com/help/publisher/182222309230722 [https://perma.cc/2WWV-FLZX].

II. PROBLEMS WITH MACHINE LEARNING-BASED CONTENT MODERATION

There are two main reasons why automated content moderation on social media is not an effective solution for dealing with misinformation in that space. The first is an empirical problem. Technology cannot effectively deal with the scale of information that exists on social media. The second is an ethical problem. Social media companies are ill-equipped, from an ethical perspective, to solve the misinformation problem with technology alone.

A. Empirical Problems with Automated Content Moderation on Social Media

Classification models are very effective for certain tasks, such as identifying easy-to-label characteristics.[12] In certain situations they can even perform better than humans.

For a complicated task like identifying misinformation, though, it is unlikely that we can expect a high rate of successful classification. After all, even humans that study misinformation for a living disagree quite a bit about what misinformation is, what its characteristic attributes are, and how to identify it.[13] Feeding information like the source of a post, the words it includes, and responses to the post into a machine learning algorithm is likely insufficient to effectively identify whether the content is true or false, even if we could agree on what truth—or the absence of truth in the form of misinformation—is in the first place.

In general, such algorithms are considered effective if they successfully classify at a level substantially greater than chance. For example, asking an algorithm to tag chat conversations as either emotional or non-emotional gives it two possible options. Flipping a coin or choosing randomly between the two (or simply always choosing one option) would be expected to produce 50% accuracy. A study that used machine learning to perform this task resulted in 90% accuracy.[14] The key here is not simply that 90% is relatively high, but that it is high in relation to 50%, which is the baseline expectation.

Machine learning classifiers specifically trained to identify misinformation on social media vary in their success rates. Two recent

---

[12] *See, e.g.,* Fariba Karimi et al, *Inferring Gender From Names on the Web: A Comparative Evaluation of Gender Detection Methods*, 25 PROC. INT'L CONF. COMPANION ON WORLD WIDE WEB 53–54 (Apr. 2016) (discussing the applicability of machine learning algorithms to infer an individual's gender based on their name).

[13] *See* Emily K. Vraga & Leticia Bode, *supra* note 1 at 136.

[14] Lars E. Holzman & William M. Pottenger, Classification of Emotions in Internet Chat: An Application of Machine Learning Using Speech Phonemes 50 (2003) (unpublished Technical Rep. LU-CSE-03-002, Lehigh Univ.).

examples achieved relatively high levels of success at this task—90.1% on a medical discussion board and a mean average precision rate of .95 on Twitter.[15] These are impressive numbers! But, these numbers also mean that at least 5% and possibly closer to 10% of problematic content on social media will fail to be identified using automated content moderation.

This is not to discount the value of automated identification of problematic content—certainly that has to be part of the solution. But when dealing with the scale of content we're talking about on social media, even a very effective identification algorithm will fail with enormous frequency.

Social media is big. Facebook, the largest social media platform in the world, currently has 2.4 billion users–roughly one out of three people in the world.[16] YouTube and WhatsApp also have more than one billion users each.[17] Massive user bases create an enormous amount of content. To provide a few illustrative examples, 300 hours of video are uploaded every minute on YouTube.[18] Every day, 60 billion texts are sent on WhatsApp, 95 million pictures are shared on Instagram, 140 million tweets are posted, and 300 million photos are uploaded to Facebook.[19]

As an example, consider this back of the envelope calculation. There are 60 billion WhatsApp texts each day.[20] If 8.5% of those texts are misinformation,[21] this would result in 5.1 billion pieces of false content shared through WhatsApp each day. Given this scope, even an algorithm that was able to accurately identify misinformation at an astounding rate of 99.9% (which is quite unlikely, given the difficulty of doing so, and in comparison to the rates cited above), would leave up something in the neighborhood of 510 million pieces of false content on WhatsApp every day.

To the extent that we think the problem of misinformation is inherent in the nature of misinformation itself—that is, the problem is not just that there

---

[15] *See* Alexander Kinsora et al., *Creating a Labeled Dataset for Medical Misinformation in Health Forums*, 2017 PROC. IEEE INT'L CONF. ON HEALTHCARE INFORMATICS (ICHI) 456, 456; Vahed Qazvinian et al, *Rumor Has It: Identifying Misinformation in Microblogs*, 2011 PROC. CONF. ON EMPIRICAL METHODS IN NAT. LANGUAGE PROCESSING 1589, 1589.

[16] *See* Esteban Ortiz-Ospina, *The Rise of Social Media,* OUR WORLD IN DATA (Sept. 18, 2019), https://ourworldindata.org/rise-of-social-media [https://perma.cc/4WY8-6H9F] (stating that 2.4 billion of the world's 7.7 billion people use Facebook).

[17] *Id.*

[18] Dustin W. Stout, *Social Media Statistics 2020: Top Networks By the Numbers,* DUSTIN STOUT, https://dustinstout.com/social-media-statistics/ [https://perma.cc/C8SP-7U2F].

[19] *Id.*

[20] *Id.*

[21] This hypothetical assumes that the 8.5% rate of false misinformation sharing among people from the Guess et al. study is a constant rate that applies across platforms and to message traffic rather than on Facebook and to people. *See* Andrew Guess et al., *supra* note 5. This assumption is likely untrue, but it is the best data-based assumption we have available; the actual percentage of texts that are misinformation could be lower or higher.

is lots of inaccurate information, but if *any* false content is being disseminated, that necessarily undermines the information environment[22]—the inability to successfully identify all such content is particularly problematic.

And this is to say nothing of the false positive rate—the extent to which content that is not in fact problematic is identified as such. That would result in true content getting falsely identified as misinformation, and interfere with user experience and any commitment to free speech that a platform supports.

B. Ethical Problems with Automated Content Moderation on Social Media

Technical limitations aside, there remains the question as to whether we are comfortable with technology platforms deciding what should be classified as misinformation in the first place. Platforms take a variety of policy approaches to this question, and sometimes even for different specific issues.

For example, Pinterest has decided there is sufficient evidence about the safety of vaccinations to make this a clear-cut issue when it comes to misinformation.[23] As a result, they still allow people to post about vaccinations but do not return any user-generated information when someone searches for information about vaccinations; rather they direct users to "reliable results about immunizations from leading public health organizations, including the World Health Organization (WHO), the Centers for Disease Control (CDC), the American Academy of Pediatrics (AAP) and the WHO-established Vaccine Safety Net (VSN)".[24] Twitter has pursued a similar approach on the question of vaccines—search results returned for a vaccine-related query also return reliable public health information (e.g., from the United States Department of Health and Human Services)—but has a much different policy when it comes to political posts.[25] For instance, when someone on Twitter (1) is a candidate for government office or current government official, (2) has more than 100,000 followers, and (3) has a verified account, they are automatically subject to a different set of rules, as compared with normal

---

[22] Indeed, the fact that people perceive misinformation on social media to be a massive problem, despite the fact that data shows only 8.5% of people share misinformation on social media, suggests that this is the case—people are concerned about misinformation even when it is not widespread. *See* Andrew Guess et al., *supra* note 5; Brett Neely, *supra* note 6.

[23] Press Release, Pinterest, Bringing Authoritative Vaccine Results to Pinterest Search (Aug. 28, 2019), https://newsroom.pinterest.com/en/post/bringing-authoritative-vaccine-results-to-pinterest-search [https://perma.cc/8RUF-AK66].

[24] *Id.*

[25] Del Harvey, *Helping you Find Reliable Public Health Information on Twitter,* TWITTER (May 10, 2019), https://blog.twitter.com/en_us/topics/company/2019/helping-you-find-reliable-public-health-information-on-twitter.html [https://perma.cc/AD7Z-P6CW].

Twitter users.[26] If such a user posts content that "may be in the public's interest" but would otherwise violate Twitter's rules (including those related to misinformation), the content is allowed to stay up, although with a notice about how it violates Twitter policy placed over the content.[27] Things get even more complicated when considering different platforms' approaches to political advertisements, which range from banning them entirely (including Pinterest, Twitter, TikTok, LinkedIn) to limiting targeting abilities (Google/YouTube) to limiting ads to those from national candidates (Reddit).[28]

This set of examples indicates just how challenging it can be to decide what counts as misinformation and when free speech or public speech or political speech should be prioritized over truthful speech. People often think of the vaccine example, but most misinformation is much more complicated than that. When it comes to vaccines, there is strong scientific consensus about safety and efficacy.[29] There are clear sources of expert information to share, and these sources even tend to be non-political and have bipartisan approval (80% of both Republicans and Democrats approve of the Centers for Disease Control, for instance).[30] The cases for which both expertise and evidence are

---

[26] *Defining Public Interest on Twitter*, TWITTER (June 27, 2019), https://blog.twitter.com/en_us/topics/company/2019/publicinterest.html [https://perma.cc/X4CN-9N8Y].

[27] *Id.*

[28] *See Factbox: How Social Media Services Handle Political Ads,* REUTERS (Jan. 9, 2020), https://www.reuters.com/article/us-usa-election-advertising-factbox/factbox-how-social-media-services-handle-political-ads-idUSKBN1Z824O [https://perma.cc/FA2X-V653]; *Platform Advertising*, CITAP, https://citapdigitalpolitics.com/?page_id=33 [https://perma.cc/RG6A-7ANQ].

[29] *See generally* German Lopez, *Does the Scientific Community Support Vaccination?,* VOX, (Aug. 25, 2016), https://www.vox.com/2018/8/21/17588092/vaccines-science-community-evidence [https://perma.cc/6LEL-S2XF]; CTRS. FOR DISEASE CONTROL & PREVENTION, *Vaccine Safety*, https://www.cdc.gov/Features/VaccineSafety/ [https://perma.cc/XF6H-XCHY]. *See, e.g.*, W.E.P. Beyer et al., *Immunogenicity and Safety of Inactivated Influenza Vaccines in Primed Populations: A Systematic Literature Review and Meta-Analysis*, 29 VACCINE 5785 (2011), https://www.sciencedirect.com/science/article/pii/S0264410X11007602?via%3Dihub [https://perma.cc/DC2A-UH8B] (regarding flu vaccine); Beibei Lu et al., *Efficacy and Safety of Prophylactic Vaccines against Cervical HPV Infection and Diseases Among Women: A Systematic Review & Meta-Analysis*, 11 BMC INFECTIOUS DISEASES at 1 (2011), https://link.springer.com/article/10.1186/1471-2334-11-13 [https://perma.cc/D9SH-9BPH] (regarding HPV vaccine); Shu-Juan Ma et al., *Combination Measles-Mumps-Rubella-Varicella Vaccine in Healthy Children: A Systematic Review and Meta-Analysis of Immunogenicity and Safety*, 94 MED. at 1 (2015), https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4915870/ [https://perma.cc/94LV-UJ6P] (regarding MMR vaccine).

[30] *Public Expresses Favorable Views of a Number of Federal Agencies*, PEW RESEARCH CTR.

this clear are unfortunately rare.[31] Indeed, for many issues, we cannot even agree who the experts are in the first place.

Because most issues lack clear guidance from experts and evidence, deciding truth is not so straightforward. And the less straightforward the issue, the less likely automated content moderation is to succeed.

III. TOWARDS A MISINFORMATION SOLUTION

Both empirical challenges and ethical issues suggest that a technology-based solution to misinformation on social media is incomplete at best. The question remains as to what other potential solutions might help solve this problem.

One option that holds a great deal of promise is person-to-person correction on social media. When someone shares misinformation, it is either seen by some portion of their social network on that platform or, depending on the platform and the individual user's setting, seen publicly. The people viewing the misinformation therefore have an opportunity to address it. Specifically, they can reply, identifying it as misinformation and correcting the original poster accordingly. This correction can originate with ordinary users or with experts or expert organizations in the area from which the misinformation comes.

The main benefit of this correction is not actually to the original poster of the misinformation (although this type of correction also tends to reduce misperceptions).[32] Indeed, that person may well feel threatened by being corrected in a semi-public space. Rather, the greatest benefit is what my research refers to as *observational correction*.[33] All of the people on the social media platform who witness both the misinformation and the correction are likely to be affected.

(Oct. 1, 2019), https://www.people-press.org/2019/10/01/public-expresses-favorable-views-of-a-number-of-federal-agencies/ [https://perma.cc/4A6J-RGNS].

[31] *See generally* Emily K. Vraga & Leticia Bode, *Defining Misinformation and Understanding its Bounded Nature: Using Expertise and Evidence for Describing Misinformation*, 37 POL. COMM. 136 (2020).

[32] *See* ETHAN PORTER & THOMAS J. WOOD, FALSE ALARM: THE TRUTH ABOUT POLITICAL MISTRUTHS IN THE TRUMP ERA 1 (2019); Emily A. Thorson, *Comparing Approaches to Journalistic Fact Checking*, *in* MISINFORMATION AND MASS AUDIENCES 249 (B. G. Southwell et al., eds., 2018).

[33] *See generally* Emily K. Vraga & Leticia Bode, *Using Expert Sources to Correct Health Misinformation in Social Media*, 39 SCI. COMM. 621 (2017) [hereinafter Vraga & Bode, *Using Expert Sources*].

Increasingly, research bears out this promise. People who observe someone else being corrected reduce their own misperceptions on the issue.[34] This is true across a variety of social media platforms—including Facebook,[35] Twitter,[36] WhatsApp,[37] and video platforms.[38]

Person-to-person correction also works whether the correction comes via a fact checker via the platform itself,[39] via an expert in the field,[40] or via another social media user.[41] This gives a great deal of flexibility in terms of who can have impact. Given that the current usership of Facebook is 2.4 billion people—nearly a third of the world's population—the potential exists to mobilize a veritable army of correctors.[42]

And again, an individual correction is observed by many *other* people, amplifying its effect. The nature of social media networks—for instance, the average Facebook user has 338 friends,[43] and the average Twitter user has 707 followers[44]—means a single correction of misinformation may be seen by hundreds of other users.

---

[34] *See generally* Leticia Bode & Emily K. Vraga, *In Related News, That Was Wrong: The Correction of Misinformation Through Related Stories Functionality in Social Media,* 65 J. OF COMM. 610 (2015) [hereinafter Bode & Vraga, *In Related News, That Was Wrong*]; Ciarra N. Smith & Holli H. Seitz, *Correcting Misinformation About Neuroscience via Social Media.*, 41 SCI. COMM 790 (2019).

[35] Leticia Bode & Emily K. Vraga, *See Something, Say Something: Correction of Global Health Misinformation on Social Media*, 33 HEALTH COMM. 1131 (2018) [hereinafter Bode & Vraga, *See Something, Say Something*]; Bode & Vraga, *In Related News, That Was Wrong*, *supra* note 34.

[36] Drew B. Margolin et al., *Political Fact-Checking on Twitter: When Do Corrections Have an Effect?* 35 POLITICAL COMMC'N 196, 196–219(2018); Vraga & Bode, *Using Expert Sources*, *supra* note 33; Bode & Vraga, *See Something, Say Something*, *supra* note 35.

[37] Sumitra Badrinathan et al., "I Don't Think That's True, Bro!" An Experiment on Fact-checking WhatsApp Rumors in India 31 (Jan. 22, 2020) (unpublished working paper), https://sumitrabadrinathan.github.io/Assets/Paper_Whatsapp.pdf [https://perma.cc/S9J2-UJEU].

[38] *See generally* Emily K. Vraga et al., The Effects of a News Literacy Video and Real-Time Corrections to Video Misinformation Correction on Health Misperceptions Related to Sunscreen and Skin Cancer (2020) (unpublished manuscript) (on file with author).

[39] Bode & Vraga, *In Related News, That Was Wrong*, *supra* note 34.

[40] Vraga & Bode, *Using Expert Sources*, *supra* note 33.

[41] Bode & Vraga, *See Something, Say Something*, *supra* note 35; Margolin, Hannack & Weber, *supra* 36.

[42] Ortiz-Ospina, *supra* note 16.

[43] Aaron Smith, *What People Like and Dislike about Facebook.* PEW RESEARCH CTR.: FACT TANK (2014), https://www.pewresearch.org/fact-tank/2014/02/03/what-people-like-dislike-about-facebook/ [https://perma.cc/2TYN-NHQ7].

[44] Kit Smith, *60 Incredible and Interesting Twitter Stats and Statistics,* BRANDWATCH (Jan. 2, 2020) https://www.brandwatch.com/blog/twitter-stats-and-statistics/ [https://perma.cc/RTQ6-ZRT8].

Evidence shows and numbers suggest great promise in mobilizing experts and users to correct one another. But the problem remains as to how to convince them to actually do so. This convincing will require a shift in public opinion.

Right now, people are much more likely to blame institutions like media and technology companies than they are to hold the public responsible for solving the misinformation problem. Two recent polls[45] found people are most likely to task the media or journalists with solving this problem.[46] Others sharing this blame include the government[47] and technology companies.[48] Yet only 20% of those polled by Pew and 12% of those polled by Marist thought the *public* should shoulder the blame for solving the problem of misinformation on social media.[49] The public will need to take greater ownership of the problem in order to normalize peer-to-peer correction efforts, and encourage more people to engage in them.

In addition, experts and users alike seem hesitant to engage in these person-to-person correction efforts. Experts may use social media to disseminate true information or even to debunk well-known misinformation. But they generally avoid engaging one-on-one with individuals in the way that research suggests might be most effective for other social media users to witness.[50]

Similarly, social media users report being somewhat hesitant to engage with other users, attempting to avoid the infamous "Facebook fight."[51] Despite this reticence, correction does occur.[52] A recent report from the United Kingdom shows that nearly three quarters of those who admit to sharing

---

[45] One was conducted by Pew and the other by Marist, PBS NewsHour, and NPR. Amy Mitchell et. al, *Many Americans Say Made-Up News Is a Critical Problem That Needs To Be Fixed*, PEW RESEARCH CTR. (JUN. 5, 2019), https://www.journalism.org/2019/06/05/many-americans-say-made-up-news-is-a-critical-problem-that-needs-to-be-fixed/ [https://perma.cc/3CUA-TSM4]; Brett Neely, *NPR Poll: Majority Of Americans Believe Trump Encourages Election Interference,* NPR (Jan. 21, 2020), https://www.npr.org/2020/01/21/797101409/npr-poll-majority-of-americans-believe-trump-encourages-election-interference [http://perma.cc/FHV2-94ZV]. The next three footnotes reference these two studies.

[46] 53% of people say journalists according to the Pew study; 39% say media according to the Marist study.

[47] 12% of people according to the Pew study; 15% according to the Marist study.

[48] 9% of people according to the Pew study; 18% according to the Marist study.

[49] Mitchell et al, *supra* note 46; Neely, *supra* note 46.

[50] Vraga & Bode, *Using Expert Sources*, *supra* note 33.

[51] Emily K. Vraga et al., *How Individual Sensitivities to Disagreement Shape Youth Political Expression on Facebook*, 45 COMPUTERS IN HUM. BEHAV. 281, 281 (2005).

[52] Ahmer Arif et al., *A Closer Look at the Self-Correcting Crowd: Examining Corrections in Online Rumors*, 2017 PROC. CONF. ON COMPUTER-SUPPORTED COOPERATIVE WORK & SOC. COMPUTING 155, 155 (2017).

"exaggerated or made up" news also report that someone corrected them for doing so, and 21% of users report engaging in such correction efforts themselves.[53] Notably, users seem to be relatively good at recognizing misinformation: only 5% of users reported that they did *not* share misinformation, but were nonetheless "corrected" by someone else for sharing accurate content.[54]

Several elements have been identified that seem to motivate people to act when they witness misinformation being shared.[55] First, there is the perceived locus of responsibility—that is, who the user thinks is responsible for the misinformation and how they view themselves as part of the information environment. This ties back to the point about public opinion—who you view as responsible for both creating and solving the problem of misinformation affects whether you are willing to act when confronted with known misinformation. Second is the corrective objective, which asks, roughly, who or what do I hope to correct? If the user thinks about the broader information space, rather than the individual they are correcting, they may be more likely to act. Related to this is the imagined audience.[56] When considering whether to correct, does a user think about the effect it will have on the person they are correcting, or on the broader audience that might view and benefit from that correction?

This model of decision-making when it comes to correction[57] offers guidance for how we might motivate people to be more willing to act when they encounter misinformation on social media. Specifically, an intervention reminding people of the broader audience for the misinformation post, and for any correction of it, might encourage more users to correct one another when appropriate.

Other interventions might be necessary for public health organizations to get more involved in this space. Research must consider the barriers to such actions, and incentives for engaging in them, when it comes to public health organizations.

---

[53] ANDREW CHADWICK & CRISTIAN VACCARI, ONLINE CIVIC CULTURE CENTRE, NEWS SHARING ON UK SOCIAL MEDIA: MISINFORMATION, DISINFORMATION, AND CORRECTION 5, 24 (2019).
[54] *Id.*
[55] Arif et al., *supra* note 52.
[56] Eden Litt, *Knock, knock. Who's There? The Imagined Audience*, 56 J. BROADCASTING & ELECTRONIC MEDIA 330, 330 (2012).
[57] Arif et al., *supra* note 52.

IV. CONCLUSION

Although technological solutions alone cannot and should not solve the problem of misinformation on social media, user- and expert-driven correction offers a data-supported means of addressing the issue that relies on individual understandings of the truth. Journalists, fact-checkers, and public health organizations might think about how they can share content that makes it easy to make use of such corrections, and technology companies and media literacy interventions should consider how to motivate people to engage with one another on the veracity of shared content on social media.