Mitigating Conspiracy Thinking with Non-Expert Crowds

Jordan S. Huffaker
Computer Science & Engineering
University of Michigan, Ann Arbor
jhuffak@umich.edu

Walter S. Lasecki
Computer Science & Engineering
School of Information
University of Michigan, Ann Arbor
wlasecki@umich.edu

Mark S. Ackerman
Computer Science & Engineering
School of Information
University of Michigan, Ann Arbor
ackerm@umich.edu

ABSTRACT

As people have become increasingly dependent on the internet for their information needs, they remain vulnerable to interpretations of that information that fit their preexisting attitudes. Adversarial actors have exploited this vulnerability by peddling narratives rationalized with *conspiracy thinking*, or the logical dysjunctions made while theorizing a conspiracy, creating ambiguous information, or creating misinformation. We are building a system —called ConTrails— to combat conspiracy thinking by both detecting it and by performing interventions on potentially vulnerable information seekers. In order to do this, our system uses the intelligence of a non-expert crowd coordinated with novel crowdsourcing workflows. Preliminary evidence suggests the potential for our system to improve the

CHI '19, June 03–05, 2018, Glasgow, UK

ACM ISBN 978-1-4503-9999-9/18/06…$15.00
https://doi.org/10.1145/1122445.1122456

Table 1: Example of Conspiracy Thinking

Example
NASA faked the moon landing because they were under pressure to demonstrate dominance over the Soviets.
Explanation
There is an implicit logical jump in that a coordinated effort to fake the moon landing would require potentially 10,000 NASA employees who played a role in the operation to keep the secret.

Table 2: Two Dyslogic Types

Bandwagon
An argument that argues for the correctness of an idea based on its popularity among a group of people.
Appeal to Emotion
An argument that uses emotion rather than facts as evidence to support a point.

reasoning capabilities of information seekers by aiding them to avoid problematic interpretations of information, thereby improving critical thinking and the engagement of readers.

CCS CONCEPTS

• **Human-centered computing** → **Computer supported cooperative work**.

KEYWORDS

misinformation, collective reasoning, crowdsourcing, conspiracy thinking

ACM Reference Format:
Jordan S. Huffaker, Walter S. Lasecki, and Mark S. Ackerman. 2019. Mitigating Conspiracy Thinking with Non-Expert Crowds. In *CHI '19: ACM Workshop on HCI for Accurate, Impartial, and Transparent Journalism: Challenges and Solutions, May 04–09, 2019, Glasgow, UK*. ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/1122445.1122456

INTRODUCTION

Information is "the currency of citizenship" to democratic institutions that rely on collective decision making processes to determine policy [2]. Yet, belief in incorrect information —or *misinformation*— is rampant in America [1]. This is especially problematic considering that such information strongly shapes citizens' attitudes and is difficult to correct once it is believed [8]. It is important, therefore, to intervene against misinformation before it is incorporated into citizens' belief systems.

As Marwick and Lewis catalog, many adversarial actors are working to peddle misinformation in online settings [10]. This is problematic because information seekers are vulnerable to believing misinformation; we highlight two of the underlying reasons why. The first vulnerability is that when presented with a set of facts, people tend to believe the facts that best align with their preexisting attitudes and reject those that misalign [11]. For systems builders, the implication of such a vulnerability is to build tools that support the assessment of facts by evaluating their *credibility*. However, there is a second reason for people's vulnerability to misinformation: people tend to form interpretations of ambiguous information that best aligns with their preexisting attitudes [4]. The implication for this second vulnerability is that systems builders need to build tools to support the *interpretation process* that people go through when incorporating information into their beliefs. Existing systems research has focused on the first vulnerability, but the second vulnerability has yet to be explored.

It is particularly important we address the second vulnerability because media manipulators have learned to exploit it by using —what Jane and Fleming coined [6]— **conspiracy thinking**. Conspiracy thinking differs from conspiracy theories: while conspiracy theories are topics of paranoia, conspiracy thinking is the problematic logic that occurs while theorizing (see Table 1). In our work, we follow Jane and Fleming in defining conspiracy thinking to broadly include logical errors (i.e., logical fallacies),

cognitive biases, and highly unlikely statements. We will additionally use the term *dyslogic* to refer to the specific ways in which arguments go wrong logically (see Table 2). In the work presented here, our goal has been to develop a system that can detect dyslogics and use that information to perform effective interventions for information seekers who might be edging toward conspiracy thinking.

Table 3: Implicitly and Explicitly Including Components of Bandwagon

Explicit Group
The publics' concerns about vaccine safety are on the rise.
Implicit Group
Concerns about vaccine safety are on the rise.

Table 4: Cultural References in Writing

With References
*A **migrant caravan** is on its way to our border. There are some **bad hombres** mixed in with them.*
Without References
A group of refugees is on its way to our border. There are some criminals mixed in with them.

BACKGROUND

An important way that authors who use conspiracy thinking evade detection is by using ambiguity. In this section, we discuss two ways in which they do that: first, how authors commonly omit information to leave room for the reader to fill it in with their own ideas; and second, how authors make use of the implicit, social contexts of references to invoke meanings.

First, authors can strategically leave out information to obscure their intentions. By using this strategy authors can set up a dyslogic in their writing, but leave some components up to the reader's interpretation. Effectively, this method communicates a dyslogic while evading a reader's awareness of it. Table 3 shows an example of how the presence of the dyslogic *bandwagon* can become ambiguous by making it non-obvious who holds the popular belief. In the example, does the "Concerns" refer to the opinion of a group of medical experts or to the average citizen? Different readers will answer differently, including knowledgeable readers [13]. Observing these implicit cues may help the reader understand what the author is doing, but readers who agree with the conclusion of the argument are less likely to put in the cognitive effort to spot them [7].

Second, using the implicit and social context of references is another common technique authors use to create effective ambiguity in their misinformation. Unfortunately, the commonality of references makes them a frequent tool used in dyslogics. For example, consider Table 4 that shows how the phrase "migrant caravan" is used to frame refugees as outsiders. Additionally, the example compounds this feeling with another phrase, "bad hombres", to magnify feelings of outsiderness and villainy. References are often ephemeral and have deep contextual roots (e.g., migrant caravan brings up several impressions including aimless wandering people, outsider Middle Easterners, and possibly militant convoys). Due to this observation, it is unlikely that computational approaches will be effective at detecting conspiracy thinking in the wild (e.g., IBM's Tone Analyzer fails to detect fear in either example in Table 4). An effective detector will need to use people to resolve these references, since they can understand the implicit social contexts used as cues.

SYSTEM

In this section we will describe our system, ConTrails, for detecting conspiracy thinking and performing interventions. We assume that the baseline for information seekers is that while they read an article, they use their internal discernment to distinguish dyslogics from perfectly fine logic. Our system uses hybrid intelligence [12] to analyze news articles and detect dyslogics at the paragraph-level (see

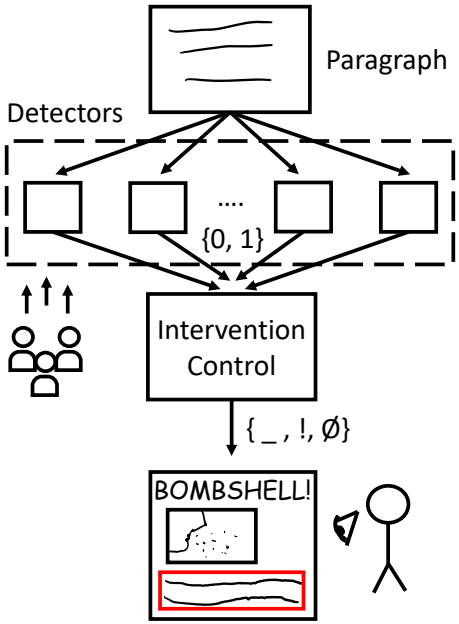

Figure 1: System for Detecting Dyslogics

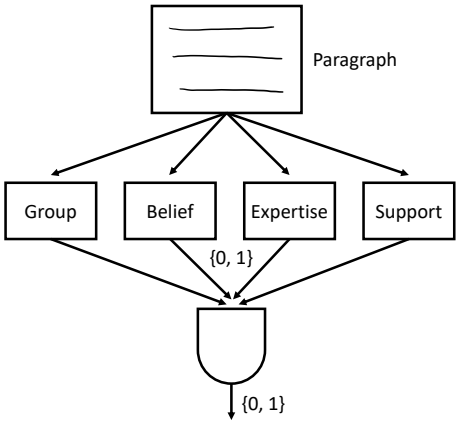

Figure 2: Bandwagon Frame Elements

Figure 1). We then use that information to compute an appropriate intervention, whether that be to do nothing, to "proceed with caution" or to block out the paragraph entirely.

For our experiments, we obtain human intelligence by recruiting workers from Amazon Mechanical Turk (AMT) with LegionTools [5, 9]. AMT is an ideal platform for worker recruitment because workers can have a vast diversity of viewpoints on any particular topic, however, preliminary evidence suggests that using such workers introduces an important challenge: they are prone to believe conspiracy thinking arguments. This is problematic because it leads to inconsistency in their detection of dyslogics. To deal with this problem, we are developing approaches for decomposing the high-level identification task into a series of subtasks that they can perform accurately.

One approach we have developed uses frame semantics [3] to decompose a dyslogic into a series of frame elements. For example, Figure 2 shows how the dyslogic bandwagon can be broken into frame elements such as a group, a belief the group holds, whether the group has expertise in the belief area, and whether the paragraph supports the belief. Each frame element is treated as a separate detector. If all fire, the paragraph is considered to contain bandwagon. While this approach is effective for dyslogics like bandwagon, some dyslogics contain frame elements that are still too difficult for the crowd to consistently detect. For example, the dyslogic *appeal to emotion* requires that the crowd assess the amount of dramatic language used that is extraneous to the information conveyed in the paragraph. We are developing an iterative approach for dyslogics such as this one.

DISCUSSION AND CONCLUSION

In this paper, we have discussed a challenging and important problem that threatens the media ecosystem: misinformation. In our discussion of the problem, we argued that conspiracy thinking is a critical reason for the vulnerability of people to misinformation, and we described a system that detects and highlights some aspects of conspiracy thinking —which we called dyslogics— for information seekers. We addressed the challenges presented in developing our system, and we showed the need for both hybrid intelligence and novel workflows of task decomposition. Succeeding in this endeavor will have implications for information sharing, journalism, HCI, and collective reasoning.

Up to this point, HCI approaches for countering misinformation have targeted the *credibility* of information, assuming that people will come to sensible interpretations when presented with the truth. This assumption has not panned out [4]. In contrast, our approach seeks to aid people in the *interpretation process* by helping them avoid problematic logic. This approach removes the need for "arbiters of truth" and instead guides people toward sensible interpretations of information by using the collective reasoning of a non-expert crowd. Future work might explore how news organizations can use article readers as crowd members in order to drive engagement and accountability.

REFERENCES

[1] Adam J Berinsky. 2018. Telling the truth about believing the lies? Evidence for the limited prevalence of expressive survey responding. *The Journal of Politics* 80, 1 (2018), 211–224.

[2] Michael X Delli Carpini and Scott Keeter. 1996. *What Americans know about politics and why it matters*. Yale University Press.

[3] Charles J Fillmore et al. 1982. Frame semantics. *Cognitive linguistics: Basic readings* (1982), 373–400.

[4] Brian J Gaines, James H Kuklinski, Paul J Quirk, Buddy Peyton, and Jay Verkuilen. 2007. Same facts, different interpretations: Partisan motivation and opinion on Iraq. *The Journal of Politics* 69, 4 (2007), 957–974.

[5] Mitchell Gordon, Jeffrey P Bigham, and Walter S Lasecki. 2015. LegionTools: a toolkit+ UI for recruiting and routing crowds to synchronous real-time tasks. In *Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. ACM, 81–82.

[6] Emma A Jane and Chris Fleming. 2014. *Modern Conspiracy: The Importance of Being Paranoid*. Bloomsbury Publishing USA.

[7] Dan M Kahan, Ellen Peters, Erica Cantrell Dawson, and Paul Slovic. 2017. Motivated numeracy and enlightened self-government. *Behavioural Public Policy* 1, 1 (2017), 54–86.

[8] James H Kuklinski, Paul J Quirk, Jennifer Jerit, David Schwieder, and Robert F Rich. 2000. Misinformation and the currency of democratic citizenship. *Journal of Politics* 62, 3 (2000), 790–816.

[9] Walter S Lasecki, Mitchell Gordon, Danai Koutra, Malte F Jung, Steven P Dow, and Jeffrey P Bigham. 2014. Glance: Rapidly coding behavioral video with the crowd. In *Proceedings of the 27th annual ACM symposium on User interface software and technology*. ACM, 551–562.

[10] Alice Marwick and Rebecca Lewis. 2017. Media manipulation and disinformation online. *New York: Data & Society Research Institute* (2017).

[11] Brendan Nyhan and Jason Reifler. 2010. When corrections fail: The persistence of political misperceptions. *Political Behavior* 32, 2 (2010), 303–330.

[12] Alexander J Quinn and Benjamin B Bederson. 2011. Human computation: a survey and taxonomy of a growing field. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1403–1412.

[13] Amy X Zhang, Aditya Ranganathan, Sarah Emlen Metz, Scott Appling, Connie Moon Sehat, Norman Gilmore, Nick B Adams, Emmanuel Vincent, Jennifer Lee, Martin Robbins, et al. 2018. A structured response to misinformation: Defining and annotating credibility indicators in news articles. In *Companion of the The Web Conference 2018 on The Web Conference 2018*. International World Wide Web Conferences Steering Committee, 603–612.