Hacking the discourse: Leaked emails, transparency, and disinformation [DRAFT]

Benjamin Sorensen
Stanford Project on Democracy and the Internet

Abstract

In scholarly and popular discourse, leaks have long been associated with transparency, but their potential as a source of disinformation has gone largely unexamined. The turmoil of the 2016 U.S. presidential election cast doubt on leaks' association with the truth by demonstrating the potentially disorienting impact of leaked emails on political knowledge and discourse. Nonetheless, the dynamics of this disruption remain poorly understood. Through an online experiment, I find that individuals perceive leaked emails as generally more credible than other forms of anonymously sourced political information. I contextualize this finding within the general understanding that leaks are just as pliable to the spread of doubt and disinformation as they are to the cause of transparency. Looking ahead, disclosures similar to the 2016 leaks are likely to figure prominently in American politics, particularly in the context of future elections. Whether these future leaks are committed in the name of transparency or disinformation, their impact on democratic practice will ultimately be determined by the reaction of their audience — journalists, politicians, and voters — who must decide whether and to what extent their contents can be trusted.

I. Introduction

At the turn of the decade, the Internet was seen by many as an imperfect yet pro-democratic force for good, a platform where the marketplace of ideas could reach a truly global scale, and where visions of a more liberal world order might be realized.[1] Today, it seems at best an ambivalent entity, one vulnerable to anti-democratic exploitation and in need of constant vigilance. Tucker, Theocharis, Roberts & Barberá (2017) describe this dilemma as the "double reality of the open online world — able to give a voice to the voiceless, but also bendable toward the aims of censorship and exclusion."[2]

This article adds to a growing literature on online disinformation and its effects on democratic politics by focusing on a phenomenon that has long embodied the double-edged nature of the Internet: digitally mediated leaks. Once most closely associated with the controversial promise of 21st century transparency and digital accountability, online leaks have more recently been weaponized by enemies of the truth to serve a more explicitly destabilizing role in democratic politics. Leaked emails, in particular, were a key vehicle for the spread of disinformation during the 2016 U.S. presidential election, and little has been done to prevent the return of similar disruption in the years ahead.

---

[1] E.g., Larry Diamond, "Liberation Technology," *Journal of Democracy* 21, no. 3 (2010).

[2] Joshua A. Tucker et al., "From Liberation to Turmoil: Social Media and Democracy," *Journal of Democracy* 28, no. 4 (2017).

The emails hacked from the DNC and John Podesta's personal inbox were disruptive insofar as they were assumed to be authentic, a gamble many reporters and their readers took with seemingly little hesitation. But the Clinton campaign refused to confirm or deny the emails' authenticity to the public, and without the victim's cooperation, private correspondence is difficult if not impossible to verify. In the media's rush to cover the click-driving contents of the leaks, the risk of amplifying altered or fabricated documents grew. Nonetheless, the emails took on a central role in the course of the campaign.

This study seeks to outline the threat leaked emails pose to political knowledge, and to empirically determine how such unreliable information can gain such wide acceptance by the public at large. This article is not meant to discount the truth-seeking value of leaks, but rather to frame the recent embrace of leaked emails as a normal part of political discourse as a risk with questionable benefits and serious downsides. A randomized control trial shows that reporting which centered its claims on evidence from a leaked email was found to be more credible than similar reporting based on other sources. This credibility enhancing effect demonstrates the power of source credibility to shape political knowledge in the online world, while also exposing the dangers and ethical pitfalls of reporting on leaked emails in the digital age.

Leaks and disinformation

The connection between leaks and disinformation is not obvious, and deserves some explanation. Tucker et al. (2018) provide a broad definition of disinformation that describes "the types of information that one could encounter online that could possibly lead to misperceptions about the actual state of the world."[3] Notably, this definition does not limit itself to the matter of "fake news," defined by Allcott and Gentzkow (2017) as "distorted signals uncorrelated with the truth."[4] In their analysis of the 2016 political media environment, Faris et al. (2017) downplay the significance of the kinds of egregious fake stories that gained attention during the campaign (such as a widely shared story which claimed the Pope had endorsed Donald Trump), and instead suggest that other "disinformation and propaganda from dedicated partisan sites on both sides of the political divide played a much greater role in the election."[5] Taken together, these definitions encompass all information — whether technically accurate or not — that is meant to mislead the public and turn the formation of political knowledge against the acquisition of truth. Thus the disinformation that stems from leaks exists in two main forms: outright falsehoods or "fake news," and also the distortion of knowledge that can incidentally result from a disclosure, e.g. from unfair or invasive media coverage, the emergence of conspiracy theories, or undue damage to the victim's reputation — in other words, distraction and doubt. At best, the disorienting effects of leaks may be greatly limited by focused, curated disclosures and investigative partnerships with credible reporters. The 2015 release of the Panama Papers offers an ideal example, as the internationally coordinated effort to understand and verify the leak's contents demonstrated the unprecedented truth-seeking potential of large-scale leaks while also preserving the anonymity of the leaker.

At worst, however, the relationship between leaks and "fake news" can be symbiotic. Leaks, especially those released on a large public scale, can provide broad cover for

[3] Joshua A. Tucker et al., *Social Media, Political Polarization and Political Disinformation* (San Mateo: The Hewlett Foundation), 3.

[4] Hunt Allcott and Matthew Gentzkow, "Social Media and Fake News in the 2016 Election," *Journal of Economic Perspectives* 32, no. 2 (2017): 212.

[5] Robert M. Faris et al., *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election*, (Cambridge: Berkman Klein Center, 2017): 21.

manipulation and deception, whether by the leakers themselves or by unassociated trolls. The kind of speculation about hidden bombshells and scandals that immediately follows large-scale leaks, along with the anticipation for more to be revealed as time goes on, allows all sorts of claims to enter the realm of believability. In 2016, for instance, online trolls seized the opportunity presented by the DNC and Podesta leaks to invent and disseminate fabricated emails intended to damage the Clinton campaign. In some cases, these forgeries seemed designed to echo broader narratives about the emails that were being promoted by credible reporting at major news outlets.[6] Even more disruptive was the popularity of extreme, conspiratorial interpretations of emails that promoted unfounded yet damaging claims about the Clinton campaign, going so far as to suggest that the leaks exposed a secret pedophile ring as well as Clinton's desire to assassinate WikiLeaks founder Julian Assange.[7]

While such *post hoc* fabrications and outlandish claims could be checked against the public contents of the leak, manipulation that occurs before publication can prove more difficult to detect. Leakers might fabricate or omit key details in a document before its release, planting the seeds for false narratives to take hold naturally through press coverage and social media discussion. This tactic appears to have been attempted by Guccifer 2.0, the alleged hacker behind the 2016 DNC leaks,[8] albeit sloppily enough to allow for eventual detection. But according to Hulcoop et al. (2017), this same strategy — which they refer to as the use of "tainted leaks" — was practiced by more cunning Russian hackers against several European public figures in 2016, and the full extent of their deception was only revealed when some of the victims allowed the researchers to cross-examine their own copies of the hacked materials.[9] Similar third-party accountability is only possible insofar as a victim is willing to corroborate the contents of the leak. It stands to reason that elected officials and those seeking office will be less likely to submit their inboxes to cross-examination,[10] meaning even a single manipulation, hidden among thousands of otherwise genuine documents, could suffice to sow blatant disinformation.

In contrast to the spread of outright "fake news," distortions to media coverage and public opinion do not require strategic deception on the part of leakers or trolls. The case of "Climategate" illustrates how even disclosures of authentic, verifiable information can be used to fuel disinformation. In 2009, unknown hackers released a cache of emails exchanged between climate researchers at the University of East Anglia and peers in their field. Climate skeptics in the media, in politics, and online began to cite certain emails as evidence to support their theories that the scientific consensus on climate change was a hoax orchestrated by biased researchers. In fact, the passages cited were simply being misunderstood, willfully or otherwise, and nine independent reviews and investigations ultimately cleared the scientists involved of any wrongdoing. Removed from the context of highly specialized, private discussions between

---

[6] E.g., as *The New York Times* and other outlets reported on what appeared to be transcripts of closed-door Clinton speeches found in the Podesta leak, fabricated passages of those same speeches began to circulate on social media.

[7] Alice Marwick and Rebecca Lewis, *Media Manipulation and Disinformation Online*, (Data & Society Research Institute, 2017): 2.

[8] Max Greenwood, "Russia-linked hacker edited DNC document to call it 'confidential': report," *The Hill* (Washington, D.C.), November 3rd, 2017. https://thehill.com/policy/cybersecurity/358662-russia-linked-hacker-edited-leaked-documents-report

[9] Adam Hulcoop et al., *Tainted Leaks: Disinformation and Phishing With a Russian Nexus* (Toronto: The Citizen Lab).

[10] The Clinton campaign adopted a general policy of neither confirming nor denying the authenticity of any leaked materials on the record, although some representatives cast general doubt on the disclosures, e.g. Vice Presidential candidate Tim Kaine's claim that "there's a capacity for much of the information in [the disclosures] to be wrong." (*Meet the Press,* 10/23/2016).

research colleagues, the emails that received the most attention took on a life of their own once they reached the public square. They also — perhaps not coincidentally — served as a distraction from the United Nations' 2009 Climate Change Conference in Copenhagen.[11] Despite efforts to correct the false narrative that had taken hold, the reputational damage suffered by the scientific community endured,[12] and skepticism stemming from the incident remains prevalent to this day.[13] Even without evidence, the suggestions of doubt and conspiracy that the leaks engendered were enough to destabilize Americans' knowledge and acceptance of climate change. Climategate showed how an invasive public accounting of private correspondence can prove damaging, even when it fails to reveal any substantive scandal.

A similar dynamic seems to have occurred in the discourse following the DNC and Podesta leaks. Both leaks quickly came to dominate the news cycle, arriving as potent distractions during the Democratic National Convention and upon the revelation of Donald Trump's crass *Access Hollywood* remarks about women, respectively. Since the 2016 election, a handful of journalists have reflected on their role in what has since become recognized as a damaging campaign of interference waged by a hostile foreign power.[14] In its discussion of WikiLeaks' election year actions and the media's uncritical response, a Pulitzer Prize winning report written for *The New York Times* offered a blunt postscript to the affair:

> WikiLeaks' best defense may be the conduct of the mainstream American media. Every major publication, including The Times, published multiple stories citing the D.N.C. and Podesta emails posted by WikiLeaks, becoming a de facto instrument of Russian intelligence.[15]

In retrospect, regardless of whether the contents of John Podesta's inbox released by WikiLeaks were genuine — and it appears that much of it was — it seems that in the frenzy to break stories, many journalists missed the forest for the trees. While a handful of reporters wrote extensively on the possibility that the hacks were a form of foreign meddling, in most coverage of the leaks, this detail was an afterthought to the invasive deconstruction of the Clinton campaign — an act of disruption which took place in plain sight.

The rash of online disinformation that has disrupted democratic politics in recent years has been studied extensively. Bots have been tracked,[16] tweets have been catalogued,[17] and social networks have been mapped,[18] all to reveal what many already suspected: that disinformation is rampant, decentralized, and still a threat. What has been somewhat lacking, by comparison, is a concerted effort to not only understand the extent of online disinformation and the role it plays in our democracy, but also how said disinformation *works*. This study is predicated on the notion

---

[11] Brian Winter, "Scientist: Leaked climate emails a distraction," *USA Today* (McLean, VA), November 24, 2009. https://usatoday30.usatoday.com/news/world/2009-11-24-copenhagen-united-nations-emails_N.htm

[12] Raghu Garud et al., "Boundaries, breaches, and bridges: The case of Climategate," *Research Policy* 43, no. 1 (2014): 60.

[13] Christian Harlos et al., "No evidence of publication bias in climate change science," *Climactic Change* 140: no. 3-4 (2017): 380.

[14] Amy Chozick, "'They Were Never Going to Let Me Be President'," *The New York Times* (New York, NY), April 20, 2018. https://www.nytimes.com/2018/04/20/sunday-review/hillary-clinton-chasing-hillary.html

[15] Lipton, Sanger and Shane, "The Perfect Weapon: How Russian Cyberpower Invaded the U.S."

[16] Emilio Ferrara, "Disinformation and Social Bot Operations in the Run Up to the 2017 French Presidential Election," *First Monday* 22 , no. 8 (2017).

[17] Faris et al., *Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election*.

[18] Chengcheng Shao et al., "Tracking and Characterizing the Competition of Fact Checking and Misinformation: Case Studies," *IEEE Access* 6 (2018).

that leaked emails are best understood in terms of their audience, rather than their perpetrators or victims. No matter how cunning a leaker or how savvy a victim's response, there is only so much either can do to influence the a leak's reception in the public square. To that end, the experiment described in this study seeks to measure how knowledge that information comes from a leaked email impacts perceptions of its credibility in the context of political reporting. By examining the perception of leaked emails at an individual-level in a randomized control trial, this experiment provides novel insight into the processes that facilitate both the formation and corruption of political knowledge.

II. Overview of experiment and results

Overview of the experiment[19]

To evaluate individual-level perceptions of leaked emails in the media, respondents were recruited through Amazon's Mechanical Turk service to answer questions and react to excerpts from a news article in a randomized control trial. The survey was publicly available on the MTurk platform until a response quota had been met, and any user with an American high school diploma was eligible to participate (this qualification, enforced by Amazon, served as a cost-effective way to target the relevant population without collecting personal information from respondents). Respondents were compensated for completion of the survey, which was designed to take just five minutes to complete.[20] A small fraction of responses were blindly excluded from the final sample in cases where survey metadata collected from Qualtrics indicated that the respondent had skipped past the reading portions of the survey at an unrealistic speed, leaving a final sample size of $n = 704$. Because participation on MTurk is self-selecting, the sampling mechanism was non-random. Samples taken from MTurk also typically skew younger, lower income, and less diverse than more nationally representative samples used in larger and more costly studies.[21] That said, while MTurk's sampling frame is by no means a perfect representation of the U.S. population, it has been shown to produce samples more representative than those typically taken by convenience methods,[22] and in recent years it has received increasing acceptance as a low-cost means of survey distribution for social science. Thus, for general practice, MTurk is considered a fair alternative to more expensive and far-reaching sampling services, and it is well-suited for individual-level, between-groups experimental trials. The survey was fielded upon approval from Stanford University's Internal Review Board.

Manipulating the features of news articles in experimental settings is a common empirical approach to measuring components of media credibility. Smith (2009), for instance, altered the content of news articles to find that reporters' use of anonymous sources did not tend to damage the credibility of their work, a finding that runs contrary to assumptions commonly held in newsrooms.[23] Sundar (1998) did much the same to find that source attribution in online news

---

[19] The full text of the survey is available in the appendix.

[20] Respondents were paid $0.75 upon completion of the survey, which took five minutes on average. This translates to roughly a $9.00/hour wage.

[21] Kevin E. Levay, Jeremy Freese, and James N. Druckman, "The Demographic and Political Composition of Mechanical Turk Samples," *SAGE Open* (January-March 2016): 11-12.

[22] A. J. Berinsky, G. A. Huber, and G. S. Lenz, "Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk," *Political Analysis* 20, no. 2 (2012): 351–368.

[23] Ron F. Smith, "Impact of Unnamed Sources on Credibility Not Certain," *Newspaper Research Journal* 28, no. 3 (2007): 8.

tended to boost respondents' ratings of an article's credibility and quality.[24] Thorson (2016)[25] and Nyhan and Reifler (2010)[26] have used similar techniques to study the effects of fact-checking and corrections in journalism. In the same vein, the article that formed the heart of this experiment described a minor political scandal in a state Senate race, detailing allegations of campaign finance violations by an incumbent candidate running for re-election, and was manipulated in order to measure differences across experimental groups. In order to limit the impact of participants' personal biases on their responses, the details of the article, including names, quotations, and organizations (except the Democratic and Republican parties) were fabricated for the sole purpose of this study. The deception was disclosed to respondents only upon completion of the survey.

Primary experiment and results

       Respondents were randomly assigned to three primary experimental groups meant to represent three different approaches to reporting on a single claim. In the control group, respondents were shown a version of the article in which the claim was attributed to an anonymous source speaking off the record. The treatment mechanism was implemented by attributing the same claim to a leaked email, and respondents were further divided into two groups for whom the presentation of the email differed. One group was presented with a quotation from the email, whereas the other was shown the quotation alongside an embedded plain text version of the email itself, presented in the style of WikiLeaks (see below). This difference in presentation helps to distinguish two common styles of reporting on leaks that may have varying effects on the perceptions of readers.

```
From: ▮▮▮▮▮▮@pagop.org
To: ▮▮▮▮▮@pagop.org
Date: 2014-08-06 19:07
Subject: Re: Kearse campaign
_____

thanks, will do.

also, re: the PAC issue, they need to be careful, because
ultimately TK will be held accountable for anything that goes
wrong. extra scrutiny not what they need rn

-RP

Sent via BlackBerry by AT&T
```

       The experiment itself consisted of three sections, the first asking basic demographic questions about age, political affiliation, and news consumption habits, and the second and third presenting portions of the fabricated article and asking respondents to record their reactions to its contents. The first part of the article described the premise of the scandal and the key figures involved, and was presented in identical form to each respondent. The second and final part of the article provided a new piece of negative information designed to be suggestive, yet not fully

[24] Sundar, "Effect of Source Attribution on Perception of Online News Stories," *Journalism & Mass Communication Quarterly* 75, no. 1 (1998): 55.

[25] Emily Thorson, "Belief Echoes: The Persistent Effects of Corrected Misinformation," *Political Communication* 33, no. 3 (2016).

[26] Brendan Nyhan and Jason Reifler, "When Corrections Fail: The persistence of political misperceptions," *Political Behavior* 32, no. 2 (2010): 303.

indicative of the incumbent's guilt, allowing for speculation on the part of the reader. How the article's author sourced this information varied by experimental group and constituted the treatment intervention. Both reading sections were followed by the same two questions prompting the respondent to use both a feeling thermometer to record sentiment and a Likert-type scale to express varying degrees of belief in the article's claims. These scales were used to measure the readers' sentiment toward the accused and their view of the credibility of the accusations, respectively.

 Levels of belief and sentiment were recorded at two points during the survey in order to approximate a within-unit measure of the outcomes of interest, measured as the difference between the level recorded before exposure to the experimental mechanism and the level after. That is, each response consists of both a baseline level and an updated level. Compared to a single-measure alternative, this within-unit, marginal effect better approximates how individuals' beliefs are affected by leaked emails — the actual outcome of interest. And from a methodological standpoint, the introduction of the marginal effect resolves some — though certainly not all — of the common pains associated with Likert-type scales. For instance, while it is difficult to ascribe a defined meaning to a single response collected via discrete, sparsely defined scales like those used in this study, the marginal effect accommodates a more precise interpretation. This is because even in cases where the magnitude of an individual response is ambiguous, the *direction* of the marginal effect offers certain interpretation. No matter the subjective value a respondent places on a given scale, how they update their belief in the second stage — positively, negatively, or not at all — can be mapped consistently, regardless of the nature of the internal scale.

 Lastly, a subtle manipulation, separate from the treatment mechanism, was incorporated into the experimental design in order to limit the possible impact of partisan motivated reasoning and respondents' political biases on the results. Partisans — respondents who indicated self-identification with either major party in the initial part of the survey — were presented versions of the article in which the accused incumbent's party affiliation was altered to correspond with the party they did *not* support. That is, self-identifying Republicans were shown allegations against a Democratic incumbent, and self-identifying Democrats were shown allegations against a Republican. Those who did not indicate a major party affiliation were arbitrarily shown allegations against either a Democrat or a Republican. While not sufficient to eliminate the possibility of partisan bias, this manipulation was intended to ensure that any effects, if present, were directionally consistent insofar as members of both parties engaged in partisan motivated reasoning. This measure also acted as a guard against the possibility of a "backfire effect," a phenomenon wherein information perceived as a correction to a partisan belief serves to strengthen, rather than correct, the original judgment or misperception.[27]

 Exposure to the full text of the leaked email positively affected respondents' assessments of the claim's credibility ($p < .03956$). There was no statistical difference between the beliefs reported by respondents in the control group and the group presented with only a quotation from the email. These results show that, as expected, the group exposed to the most explicit form of the treatment mechanism exhibited the largest effect. They also hint at the persuasive power of leaked emails as a form of easily manipulated and rarely scrutinized "evidence" in the context of political reporting.

 The group exposed to the full text of the leaked email also exhibited a negative marginal effect on reported sentiment ($p < .00245$). That is, the more explicit treatment negatively

---

[27] Nyhan and Riefler, "When Corrections Fail."

affected respondents' feelings toward the accused in the second stage of the experiment. No significant effect was observed in the control group or in the group exposed only to the quotation. As such, the marginal effect on sentiment mirrors the credibility effect, resulting in two simple and likely related claims: When reporting on claims that are unverifiable to an ordinary reader, inclusion of the full text of a leaked email has the effect of 1) enhancing the associated claim's credibility, and 2) enforcing negative feelings toward those implicated by the claim.

Secondary experiment and results

Embedded in each of the primary treatment conditions was a secondary mechanism designed to introduce reasonable doubt to the claims and measure its effect. Within both treatment groups, roughly half of the respondents were presented with a version of the article in which a spokesman for the accused "declined to comment" on the allegations, whereas the other half were presented with the same spokesman's more aggressive suggestion that the emails "were being misconstrued and may even have been altered." All respondents in the control group were presented with the same "no comment" response. This response — the secondary treatment tested in both groups — was designed to reflect a tactful form of implied denial that neither confirmed nor denied the emails' authenticity, making for a reasonable line of defense against claims that are still potentially dubious in the public eye. This tactic also emulates the Clinton campaign's response to the release of John Podesta's emails in October, 2016. Like the primary treatment conditions, these secondary treatments were assigned at random and designed to facilitate *post hoc* comparisons across groups.

This secondary experimental mechanism provided noteworthy results within the group exposed to the full text of the email, where an overall positive effect was observed. Among only those exposed to the secondary treatment within this group, there was no significant change in average belief from those in the control group. This diminished effect may be attributed to the doubt cast by the treatment mechanism, however a model directly comparing the sub-groups does not reveal a significant difference between the two. The possible countervailing effect of questioning the emails' authenticity should be the focus of future work.

III. Discussion and future work

Discussion

While the experimental results help paint a clearer picture of the distinct role leaked emails have played in past political discourse, they also pose a puzzle: why are leaked emails seen as *more* credible than sources that are ostensibly held to a higher standard of accountability? If anything, a reporter's willingness to stake a claim in their trust of an unnamed but privately known source should connote greater credibility than their reporting on unverifiable materials provided by a completely anonymous leaker. Yet, counter to this thinking, leaked emails appear to be more trusted than anonymous sources when they are made central to a claim.

One explanation is that readers are not accustomed to scrutinizing the contents of a leak for themselves, and may instinctively assume due diligence on the part of whoever is presenting the information. In the past, many high-profile leaks have been mediated by journalists at trusted media institutions, notably *The New York Times, The Guardian,* and *Der Spiegel.* In these cases, reporters were able to scrutinize and verify the relevant contents of the leaks before presenting their findings to the public. In the case of the 2016 DNC and Podesta leaks, on the other hand, the disclosures were made to the public and the media simultaneously, leading to rushed reporting and a storm of speculative narratives and suggestions of scandal. In cases like these, the assumption of thorough reporting was less likely to be fulfilled, yet that distinction may have

eluded ordinary readers. If it is indeed the case that a leak's audience instinctively leans toward trust over skepticism, even in the face of personal unverifiability, then reporters may better serve the public interest by disclosing their own uncertainties about the materials rather than simply presenting them at face value.

Leaks could also benefit from a negative credibility bias — the common instinct to place trust in negative information more readily than in positive or neutral information.[28] Leaks targeting political figures, in particular, are seen as a way to see past the veil of a candidate's polished image and into a darker reality. Even facially neutral information can generate negative connotations when framed as a well-kept secret that has been hidden from public view for a reason.[29] Modern news cycles likely reinforce this tendency, as reporters and pundits will tend to seize on the most scandalous narratives a leak has to offer. The results of the experiment seem to support this explanation, as the observed credibility effect is mirrored by the email's effect on sentiment, i.e. the absence of a significant change in sentiment corresponds with the absence of an effect on credibility and *vice versa*. It stands to reason that a negative credibility bias may be driving this relationship, although the data cannot be definitively interpreted to say as much. If this is indeed the case, though, one might expect a more negative disclosure to be seen as more credible, and vice versa — a possible subject for future work.

This study has several shortcomings. One stems from the fact that judgments of credibility are multidimensional, and their meaning and significance are bound to be shaped by the context in which they are formed. The scientific understanding of credibility and its constituent factors has advanced to the point where trust in the news media can be more precisely measured through a battery of carefully calibrated questions.[30] Similar methods have been developed to derive more robust measures of partisanship and other political attitudes, as well. The single Likert-type scales employed in this study are certainly incomprehensive by comparison.

The substance of the experiment is also worth interrogating, as some features of the fabricated article might have impacted the results in unseen ways. While the design was intended to account for partisan bias, for example, the correction relied on the assumption that respondents on both sides of the political spectrum would react with equal and thus counteracting biases. There can be no such guarantee, particularly since a politically charged topic — PAC spending — formed the basis for evaluation. In future work, it might be fruitful to not only control for the party identification of the accused, but for the political valence of the allegations, as well.

Lastly, it should be noted that the control condition — the reference to a source speaking off the record, rather than a leaked email — is not necessarily an ideal point of comparison for the treatment conditions. Ideally a leaked email's presence in the treatment could be compared to its simple absence in the control, rather than the added presence of another kind of source (in this case, the anonymous official). Since such symmetry was not feasible in the context of the fabricated article, the control was chosen to present its audience with evidence that, like a leaked email, was of ambiguous origins and facially unverifiable to an ordinary reader. In the article, neither could be taken as ground truth, yet neither could be refuted, either. As a result, a typical respondent's assessment of the claim's credibility would likely center on their assessment of the

---

[28] Benjamin E. Hilbig, "Sad, thus true: Negativity bias in judgments of truth," *Journal of Experimental Social Psychology* 45 (2009): 983.
[29] Clare Birchall, "Introduction to 'Secrecy and Transparency': The Politics of Opacity and Openness," *Theory, Culture, & Society* 28, no. 7-8 (2012): 8.
[30] Matthias Kohring and Jörg Matthes, "Trust in News Media," *Communication Research* 34, no.2 (2007).

source, and any observed difference between the groups' responses could be attributed to the subtle manipulation of the source which constituted the experimental intervention.

Conclusion

The results of this study contribute to the literature of online disinformation in two important ways. First, they reaffirm the importance of source credibility in determining how journalistic claims are perceived, and illustrate the potential for trolls and peddlers of fake news to exploit mainstream acceptance of leaked materials' trustworthiness. Second, they demonstrate the potential for even credible, mainstream reporting to unwittingly contribute to the spread of disinformation through practices that were commonplace in 2016. In summary, this study demonstrates that email leaks can engender a highly effective form of disinformation, conveying the appearance of credibility while eschewing, for the most part, typical expectations of accountability from the mainstream media and their public audience.

As has been discussed, an uncritical approach to reporting on the contents of leaked emails carries the possibility of promoting tainted materials and amplifying false narratives. The temptation to believe leaked materials on their face can also sow doubt and distraction into political discourse by creating news cycles conducted on the terms of the hacker, rather than the public interest. As *The New York Times* reporter Scott Shane writes, "The tilt of the coverage [in 2016] was decided in Moscow. By counting on American reporters to follow their usual rules, the Kremlin hacked American journalism."[31]

The ability of leaked emails to not only garner undeserved trust, but also to rapidly take hold of the public imagination, poses a potent threat to the healthy formation of political knowledge — especially given that research suggests even effectively correcting false beliefs can fail to reverse their effects on political attitudes.[32] As suggested by the results of the secondary experiment, however, there are steps reporters can take to limit the downsides of reporting on leaks, such as raising legitimate doubts about the authenticity of an email or being transparent about the possibility of fraud. Extra caution and transparency, paired with the recognition that not all leaked materials are newsworthy, may go a long way toward limiting the spread of this kind of disinformation. Still, social media trolling, fake news, and conspiracy mongering are likely to thrive in any environment where large-scale leaks of stolen political materials are normalized, and even responsible reporting on these materials can carry the risk of amplifying their disorienting effect by legitimizing their use. Leaks hosted on publicly searchable databases like WikiLeaks pose the particular threat of encouraging amateur investigators to apply their own ethics and standards of reporting to the contents of a leak. In the case that anyone can access the source materials, it should come as no surprise that anyone — with any intentions — can use them to tell a compelling story.

If nothing else, this article demonstrates that the implications leaked emails hold for the formation of American political knowledge — and consequently for democratic practice — should be taken seriously. Going forward, the most effective safeguard against such disruption might simply be greater caution: on the part of the public figures and organizations that must adopt greater security measures to prevent future hacks; on the part of the reporters who must undertake the difficult task of verifying and contextualizing disclosures within an inevitably vague understanding of their larger significance; and on the part of individuals and ordinary

---

[31] Eric Lipton, David E. Sanger, and Scott Shane, "The Perfect Weapon: How Russian Cyberpower Invaded the U.S.", *The New York Times* (New York, NY), Dec. 13, 2016. https://nytimes.com/2016/12/13/us/politics/russia-hack-election-dnc.html

[32] Thorson, "Belief Echoes."

Americans, who must learn to be ever more discerning about the information they choose to trust.

Appendix: Survey text

Background questions:
1. What is your age?
    a. 18-35
    b. 36-50
    c. 50+
    d. Prefer not to answer
2. What best describes your party affiliation?
    a. Democrat
    b. Republican
    c. Independent/Other
    d. Prefer not to answer or no preference
3. What best describes your political leaning?
    a. Very conservative
    b. Somewhat conservative
    c. Moderate/Centrist
    d. Somewhat liberal
    e. Very liberal
    f. Prefer not to answer or no preference
4. How often do you follow politics in the news?
    a. Never
    b. Rarely
    c. More than once a week
    d. Every day
5. When you follow politics, how often do you use online sources (e.g. online news, blogs, or social media)?
    a. Never
    b. Less than half the time
    c. More than half the time
    d. Always

The following is excerpted from the beginning of a 2014 news article. Please read it carefully — once you click past, you will not be able to view the article again.

The Pennsylvania [Democratic/Republican] Party has accused [Repulican/Democratic] state Sen. Tim Kearse of violating campaign rules by coordinating his re-election effort with a [conservative/liberal] super PAC.

Kearse's campaign and the group the senator is alleged to have worked with dismissed the charges as an act of political desperation.

In a complaint filed Monday with the Federal Elections Commission, the [Democrats/Republicans] contend the incumbent [Republican/Democrat] asked Judy Rawlins, the president of the Pennsylvania Victory Coalition, to direct more funds toward his senate race.

Super PACs, or independent expenditure-only committees, can raise unlimited money from corporations, labor unions, individuals, and other organizations. These funds can be used to advocate for or against political causes, including candidates for elected office. But it is illegal under federal law for super PACs to work directly with political campaigns, either through direct donations or strategic coordination.

"This is nothing more than a last ditch effort to distract voters from the real issues," Jeremy Carson, a spokesman for the Pennsylvania Victory Coalition, said in an email. "We have operated completely independently of the senator's campaign, and there is simply no truth to these allegations."

A Kearse campaign spokeswoman said much the same when reached for comment.

Based on the information you have read and relying solely on your best judgment of the circumstances described, please answer the following questions.

1. What party does state Sen. Tim Kearse belong to?
   a. Democratic Party
   b. Republican Party
   c. Unsure
2. On a scale from 0-100, how would you describe your sentiment toward State Senator Kearse? A rating between 51-100 indicates generally favorable or positive feelings, whereas a rating between 0-49 indicates generally unfavorable or negative feelings. [Feeling thermometer from 0-100]
3. Please indicate whether you agree or disagree with the following statement and how certain you feel:
   > Tim Kearse violated campaign regulations by coordinating with the Pennsylvania Victory Coalition during his race for state senate.
   [Likert-type scale from 1-7]

Below is the last part of the excerpted article you just read. Please read it carefully before answering the questions on the following page.

[CONTROL]

But critics note that Kearse and Rawlins share a long history together, having both risen through state politics as aides in the state legislature, and argue that this might open the door to illegal coordination.

Some state [Republican/Democratic] officials have expressed concern about the allegations and their possible impact on what is looking to be a close race. One state [Republican/Democratic] party official, speaking on the condition of anonymity, said "They need to be careful, because ultimately he'll be accountable for anything that goes wrong. Extra scrutiny is not what they need right now."

In a statement, a [Democratic/Republican] Party spokesman said: "The FEC must investigate the possibility of coordination between the Kearse campaign and the Pennsylvania Victory Coalition to ensure that the election is decided fairly and so that we can get back to the real issues at hand."

[TREATMENT 1: QUOTED EMAIL]

But critics note that Kearse and Rawlins share a long history together, having both risen through state politics as aides in the state legislature, and argue that this might have opened the door to illegal coordination.

Some state [Republican/Democratic] officials have expressed concern about the allegations and their possible impact on what is looking to be a close race. In emails leaked from state [Republican/Democratic] party servers, one official wrote to a colleague about his concerns: "they need to be careful, because ultimately [Kearse] will be held accountable for anything that goes wrong. extra scrutiny is not what they need [right now]."

[A party spokesman, when asked about the emails and party officials' apparent misgivings on the matter, declined to comment.] / [A party spokesman, when asked about the emails and party officials' apparent misgivings on the matter, raised the notion that the leaked documents were "being misconstrued and may even have been altered," without clarifying which documents were in question.]

In a statement, a [Democratic/Republican] Party spokesman said: "The FEC must investigate the possibility of coordination between the Kearse campaign and the Pennsylvania Victory Coalition to ensure that the election is decided fairly and so that we can get back to the real issues at hand."

[TREATMENT 2: FULL EMAIL]

But critics note that Kearse and Rawlins share a long history together, having both risen through state politics as aides in the state legislature, and argue that this might have opened the door to illegal coordination.

Some state [Republican/Democratic] officials have expressed concern about the allegations and their possible impact on what is looking to be a close race. In emails leaked from state [Republican/Democratic] party servers, one official wrote to a colleague about his concerns: "they need to be careful, because ultimately [Kearse] will be held accountable for anything that goes wrong. extra scrutiny is not what they need [right now]."

```
From: ███████@[pagop/padems].org
To: ██████@[pagop/padems].org
Date: 2014-08-06 19:07
Subject: Re: Kearse campaign
_____

thanks, will do.

also, re: the PAC issue, they need to be careful, because
ultimately TK will be held accountable for anything that goes
wrong. extra scrutiny not what they need rn

-RP

Sent via BlackBerry by AT&T
```

[A party spokesman, when asked about the emails and party officials' apparent misgivings on the matter, declined to comment.] / [A party spokesman, when asked about the emails and party officials' apparent misgivings on the matter, raised the notion that the leaked documents were "being misconstrued and may even have been altered," without clarifying which documents were in question.]

In a statement, a [Democratic/Republican] Party spokesman said: "The FEC must investigate the possibility of coordination between the Kearse campaign and the Pennsylvania Victory Coalition to ensure that the election is decided fairly and so that we can get back to the real issues at hand."

Based on the information you have read and relying solely on your best judgment of the circumstances described, please answer the following questions.

1. On a scale from 0-100, how would you describe your sentiment toward State Senator Kearse? A rating between 51-100 indicates generally favorable or positive feelings, whereas a rating between 0-49 indicates generally unfavorable or negative feelings. [Feeling thermometer from 0-100]
2. Please indicate whether you agree or disagree with the following statement and how certain you feel:
   Tim Kearse violated campaign regulations by coordinating with the Pennsylvania Victory Coalition during his race for state senate.
   [Likert-type scale from 1-7]