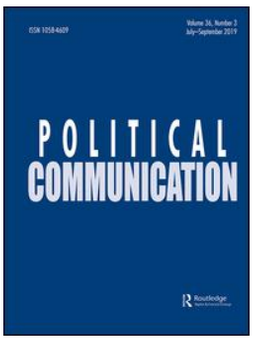

Political Communication

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/upcp20

Believing and Sharing Information by Fake Sources: An Experiment

Paul C. Bauer & Bernhard Clemm von Hohenberg

To cite this article: Paul C. Bauer & Bernhard Clemm von Hohenberg (2020): Believing and Sharing Information by Fake Sources: An Experiment, Political Communication, DOI: 10.1080/10584609.2020.1840462

To link to this article: https://doi.org/10.1080/10584609.2020.1840462

POLITICAL COMMUNICATION
https://doi.org/10.1080/10584609.2020.1840462

Routledge
Taylor & Francis Group

Believing and Sharing Information by Fake Sources: An Experiment

Paul C. Bauer [a] and Bernhard Clemm von Hohenberg [b]

[a]Mannheim Centre for European Social Research (MZES), University Mannheim, Mannheim, Germany;
[b]Department of Social and Political Sciences , European University Institute, Florence, Italy

ABSTRACT

The increasing spread of false stories ("fake news") represents one of the great challenges societies face in the 21st century. A little-understood aspect of this phenomenon and of the processing of online news in general is how sources influence whether people believe and share what they read. In contrast to the predigital era, the Internet makes it easy for anyone to imitate well-known and credible sources in name and appearance. In a preregistered survey experiment, we first investigate the effect of this contrast (real vs. fake source) and find that subjects, as expected, have a higher tendency to believe and a somewhat higher propensity to share news by real sources. We then expose subjects to a number of reports manipulated in content (congruent vs. incongruent with individuals' attitudes), which reveals our most crucial finding. As predicted, people are more likely to believe a news report by a source that has previously given them congruent information. However, this only holds if the source is fake. We further use machine learning to uncover treatment heterogeneity. Effects vary most strongly for different levels of trust in the mainstream media and having voted for the populist right.

KEYWORDS

disinformation; propaganda; fake news; misinformation; fake sources; media trust

Introduction

How citizens manage to stay informed about politics – or not – has been a long-standing concern of political science (Berelson et al., 1954; Delli Carpini & Keeter, 1996; Druckman, 2014; Lippmann, 1922). In a complex world, people must turn to others for such information. Today, the Internet's "many-to-many" structure (Tucker et al., 2017; Van Aelst et al., 2017) makes it harder to know what sources to turn to: trust in professional media has been waning in many Western democracies (Gallup, 2019; PricewaterhouseCoopers, 2018) and is increasingly related to partisanship (Arceneaux et al., 2012; Tsfati & Ariely, 2014). At the same time, the spread of misinformation is rampant (Guess et al., 2018; Vosoughi et al., 2018). How do news sources affect belief and sharing of their information in this environment?

Research on sources reaches back decades (e.g., Hovland & Weiss, 1951), but given the rapidly changing news ecology, it deserves renewed attention. In an original, preregistered survey experiment, we examine how sources influence whether people believe and intend

CONTACT Paul C. Bauer mail@paulcbauer.eu Mannheim Centre for European Social Research (MZES), University Mannheim, Mannheim 68131, Germany.

*Both authors contributed equally to this work.

Supplemental data for this article can be accessed on the publisher's website at https://doi.org/10.1080/10584609.2020.1840462

to share news reports in an online context. Previous studies on the effect of sources have mostly compared partisan outlets or politicians (Baum & Gussin, 2004; Druckman, 2001; Swire et al., 2017). In this study, we focus on a different aspect, namely the appearance of fraudulent sources. On the Internet, it takes no big effort to invent a news brand. Misinformation entrepreneurs sometimes counterfeit existing sources, as in a recent instance of misinformation on the Corona virus (Faz, 2020). More subtly, they also use professional-sounding names to feign credibility. We mimic such a situation in our experiment by making up a source with a professional-sounding name and appearance. We compare this "fake" source with a real well-known source, and test in a first step whether this contrast affects people's belief in a news report as well as their intent to share it.

Our key contribution is to highlight the ideological determinants of source credibility by manipulating the content of news reports in the ensuing experimental stages. Specifically, we randomly expose subjects to several reports that are either congruent or incongruent with their premeasured attitudes. While it is well known that people have a tendency to believe information that supports their worldview (Jerit & Barabas, 2012; Lord et al., 1979; Taber & Lodge, 2006), we show that a source that provides the "right" facts is *subsequently* more likely to be believed. This can explain, first, why trust in and exposure to media is partisan (Goldman & Mutz, 2011; Iyengar & Hahn, 2009; Pennycook & Rand, 2019c). Second, we note that this dynamic might be exploited by malevolent agents to generate trust in fake sources.

We further contribute to the literature by thoroughly examining treatment heterogeneity with machine learning methods recently developed by Athey and Imbens (2016; also Wager & Athey 2018). In contrast to conventional approaches, the causal forest approach allows to explore heterogeneity over a large covariate space without concerns about multiple testing. We find that treatment effects differ strongest across mainstream media trust, vote choice, age and political knowledge. Our findings highlight how different groups vary in their vulnerability to fake sources and one-sided reporting.

We proceed as follows: Section 2.1 discusses previous literature on source credibility and develops expectations for the contrast between an existing and a fake source. In Section 2.2, we elaborate on our prediction that congruent news reports will influence belief and sharing intentions of subsequent information. Section 3 presents the design, data, and measures. Section 4 summarizes the results. Section 5 provides a conclusion and suggestions for future research.

Theory and Hypotheses

Fake vs. Real Sources

The importance of sources in information transmission has been a central concern of psychology at least since the work of Hovland and colleagues (Hovland et al., 1953; Hovland & Weiss, 1951). Political science joined the debate about "source cues" some decades later (Mondak, 1993; Page et al., 1987; Popkin, 1991; Sniderman et al., 1991; Zaller, 1992). Both disciplines were mainly interested how sources affect opinion change (*persuasion*). Fewer studies tested how sources impacted *believing* factual information or intentions to *share* it. Only recently, with rising concern about fake news, scholars have

paid more attention to the impact of sources on these outcomes (e.g., Oeldorf-Hirsch & DeVoss, 2019; Pennycook et al., 2020; Swire et al., 2017).

Most of the literature links the effect of sources to the concept of *source credibility*: the more credible a source, the more people tend to be persuaded by its arguments or believe the information it provides. We follow this idea and assume that people hold subjective, not necessarily conscious, impressions about the credibility of a source, influencing outcomes such as belief and sharing intentions.[1] A host of experimental studies have tested which source characteristics determine these perceptions. One common example of a manipulation is the *professionalism* or expertise of a source. The seminal study by Hovland and Weiss (1951) compared the persuasiveness of a scientific journal to that of a pictorial monthly, a contrast adapted by many subsequent studies (e.g., Chebat et al., 1988; Greer, 2003; Petty et al., 1981; Sternthal et al., 1978). Another manipulation concerned the source's physical *attractiveness* or likability (e.g Chaiken, 1980; Mills & Aronson, 1965). *Similarity*, e.g., in social or political identity, between source and receiver was another effective treatment (e.g., Baum & Groeling, 2009; Kuklinski & Hurley, 1994; Kuru et al., 2017).

In the present study we are interested in a contrast that has become particularly important with the rise of the Internet and social media. First, let us note that we study the situation of people receiving news reports, by which we mean factual information that can be either true or false. People typically do not know ex ante whether a news report is true and usually do not have the time to verify the facts: It is in these situations that sources seem most relevant. Online, the variety of sources is practically unlimited. News reports may originate from real, well-known news organizations. Some of these might be widely perceived as ideologically biased or generally unreliable, but we focus here on real sources that generally enjoy high levels of trust. On the other hand, the Internet facilitates the emergence of fraudulent sources that do not actually represent a news organization. Anyone with some technical understanding can easily set up a Facebook or web page to spread news, and it is easily possible to imitate real news organizations in appearance and name.

We are interested in this contrast between real sources generally perceived as credible (henceforth *real* sources), and fake sources imitating them in name and appearance (henceforth *fake* sources). To illustrate the relevance of this contrast, consider the story of the made-up *Denver Guardian*, which posted a story on Hillary Clinton that garnered more than half a million interactions on Facebook (Berghel, 2017). The source name played on existing, respected brands like the *Guardian*, helped by a neutral website appearance. As documentations of the most successful fake news stories before the 2016 US election illustrate, such sources – next to more obviously hyperpartisan outlets – are a prominent factor in the dissemination of misinformation (e.g., Silverman, 2016).

The question, thus, is whether fake sources are believed substantially less than real sources. Recent studies have examined this contrast only implicitly, by exposing subjects to true and false news with the original source attributions, but have not disentangled source and content effects (Pennycook et al., 2020, 2019a, 2019b). Real and fake sources as defined above may differ on several dimensions. It might be that a fake source's name and appearance fail to signal that it reports true information. An alternative mechanism may be familiarity: people should be less inclined to believe sources they do not know. Indeed, in some recent studies, familiarity with a source has been found to affect belief (Epstein

et al., 2020; Pennycook & Rand, 2019c). Given the same news report provided by either a real or a fake source, we thus predict that *people have a higher tendency to believe news reports from a real source than from a fake source (H1a)*.

Internet users are not just passive receivers of news but also share them. We adapt a broad definition of sharing, which includes any dissemination of news reports via social media, but also messengers and e-mail. Researchers are only beginning to understand what makes news "shareworthy" (Barberá et al., 2015; Boczkowski & Mitchelstein, 2012; Trilling et al., 2017). It is unclear whether sources affect sharing in a similar way as they affect belief. There could be a path through belief: people share information they believe to be true. Yet, there is evidence that believing something is not a prerequisite for sharing it. People might be motivated to distribute information they know is false for political gain (Chadwick & Vaccari, 2019) or because of a "need for chaos" (Petersen et al., 2018). Alternatively, reputational concerns might play a role: Insofar as social media is about self-presentational needs (Seidman, 2013), users might be more willing to share sources with prestige, irrespective of whether they believe the content. Without elaborating on these mechanisms, recent studies have found sources to matter for sharing intentions (Pennycook et al., 2020). In the absence of more direct evidence, we predict that people *will be more likely to share news reports that come from real sources than those that come from fake sources (H1b)*. Taken together, H1a and H1b predict what we call the source effect.

Congruence & Repeated Interactions

We start from the assumption that someone's perception of a source's credibility develops over repeated interactions. For example, if credibility partly depends on whether the source objectively tells the truth, then a source should be perceived as more credible when it provided true information and less so when the information turns out to be false. However, patterns of media trust suggest that it cannot be objective truth-telling alone that makes a source subsequently credible in the eyes of receivers: various studies show that people consider media organizations that align with them ideologically more trustworthy (American Press Institute, 2017; Tsfati & Ariely, 2014). Such purely correlational observations do not tell us about the direction of causality: we do not know whether people align ideologically with sources they have always found credible or whether they find sources more credible after having been supplied with ideologically congruent content.

We here focus on the impact of congruent reporting on later credibility, and thus belief and sharing. Our argument starts from the well-established finding in psychology and political science that people evaluate information according to its *congruence* with their preexisting attitudes, values, or beliefs (Kahan et al., 2017; Lord et al., 1979; MacCoun & Paletz, 2009; Taber & Lodge, 2006). In particular, they are more likely to accept attitudinally congruent than incongruent *factual* information, irrespective of the actual truth value (Pennycook & Rand, 2019a; Washburn & Skitka, 2017).[2]

We suggest that such one-sided processing of factual information will influence later credibility perceptions of a source. Since people tend to believe (to not believe) congruent (incongruent) information because of their directional motivations, they have a *subjective* perception of the source as truth-telling (or not). This credibility perception will at least

partly be independent of whether the source is *objectively* truth-telling – assuming people do mostly not actively verify reporting. After repeated encounters with a source that provides congruent information, an individual will hence perceive the source as more credible. In consequence, the individual should be more likely to believe factual information (regardless of what it is) by that source later on. Studies conceiving of the credibility of a source as a probability that is updated in a Bayesian fashion come to similar predictions (Gentzkow & Shapiro, 2006; Koehler, 1993). To our knowledge, no experimental study has hitherto tested this prediction in the context of political news. In sum, we expect that *those who previously saw congruent (vs. incongruent) news reports by a source will have a higher tendency to believe a subsequent news report by the same source (H2a).*

As far as sharing intention is concerned, we predict a similar effect of congruence. Again, we do not claim that this effect necessarily goes through belief. It is possible that people learn which sources provide congruent content and can be shared without risk for one's reputation among ideologically close people. Motivation to share the "right" sources and fear of sharing "wrong" sources might be independent of truth judgment. Whichever the mechanism, we expect that *those who previously saw congruent (vs. incongruent) news reports by a source will be more likely to share a subsequent news report by the same source (H2b)*. We refer to these predictions as the congruence effect.

We further argue that this congruence effect will play out differently for real and fake sources. As argued above, credibility perceptions will depend on previous interactions. For a real source that people are likely to be familiar with, credibility perceptions have formed over a longer period. Such perceptions are likely to be sticky. Relatively few additional encounters should not change much for real sources. In contrast, for a fake source previously unknown to people, a few encounters that involve congruent information may change an individual's perception significantly. This suggests a normatively worrying dynamic: fraudulent actors may quickly gain an individual's trust by providing the "right" facts. In sum, we expect that the *difference in later news belief between those who saw congruent and those who saw incongruent reports is higher for the fake source than for the real source (H3a)*. We also expect the *difference in later sharing intentions between those who saw congruent and those who saw incongruent reports to be higher for the fake source than for the real source (H3b).*

Heterogeneity of Source and Congruence Effects

We are further interested in exploring treatment heterogeneity. As a review of five decades of credibility research notes, the "interaction between source credibility and demographics of recipients has not been researched or analyzed to a great extent" (Pornpitakpan, 2004, p. 263). We thus do not have clear expectations how the source effect (H1a, H1b) and the congruence effect (H2a, H2b) might vary across subpopulations and chose an exploratory approach without prespecifying any hypotheses. However, we want to point to some evidence that invites speculation about moderators.

First, *age* figures as an important factor impacting media behavior, e.g., online participation (Hargittai & Walejko, 2008; Loges & Jung, 2001). Pertinent to our hypothesis on source effects, there is mixed evidence how people of different ages evaluate different types of media outlets: some surveys suggest that younger people are less impressed of mainstream news sources (American Press Institute, 2017; PricewaterhouseCoopers, 2018),

while others find older people to be greater "media skeptics" (Gunther, 1992; Metzger et al., 2003).

The same studies also suggest an important role of *education* and *income*, variables that have also been found to correlate with the ability to distinguish professional and unprofessional websites (Fogg et al., 2001). There is also a debate whether higher education and related variables such as *political knowledge* lead to more or less bias in information processing (Kahan, 2013; Tappin et al., 2018). Accordingly, the effect of congruence on believing a source subsequently could vary along these characteristics.

A large number of surveys and studies have examined people's media trust, understood as an individual perception about how certain types of media, e.g., "the mass media" or "television", can be trusted, report fair and accurately, and tell the whole story (Flanagin & Metzger, 2000; Gallup, 2019; Kiousis, 2001; Kohring & Matthes, 2007; Tsfati, 2010). In particular, *mainstream media trust* has been found to correlate with behaviors of information processing. With regard to our study, it could be that people with high media trust are more affected by the source when judging news and deciding whether to share them. Relatedly, there is evidence that certain *media habits* such as frequency of social media use affect news processing (Allcott & Gentzkow, 2017; Greer, 2003).

Finally, *ideology* or *vote choice* have time and again proved to be related to media behavior and information processing (e.g., Iyengar & Hahn, 2009; Vallone et al., 1985). In the US, Republicans select and tend to believe different news sources than Democrats (Gallup, 2019; Pennycook & Rand, 2019c), however the relation between ideology and media evaluations is less clear in Germany. Scholars also debate the role of partisanship in bias in information seeking and processing (Ditto et al., 2018; Jost, 2017; Kahan, 2016b). It could be that one side of the political spectrum is more prone to believing sources that cater to their preexisting beliefs and attitudes.

Methods: Design, Data, Measures and Models

Sample

To test our predictions, we ran a preregistered online survey experiment between March 14 and 29, 2019. We didn't collect any data before the preregistration on March 12th, 2019 [see Appendix A.16 and https://osf.io/q2ucj]. The non-probability sample of Germans was recruited by the survey company Respondi AG. To enhance external validity, we screened respondents into the survey according to quotas on gender, age (5 categories), and state of residence. We excluded respondents who did not complete the questionnaire (125 respondents). Those who completed the questionnaire in less than half of the median time (308 respondents) were marked as non-completes by the survey company. Robustness tests show that their exclusion does not change results. The final sample included 1,980 participants. See Appendix A.5 for a comparison of our sample to the German population.

Experimental Design

The questionnaire and randomization logic was built with Qualtrics. The setup is depicted in Figure 1. Upon reading about the study's content and giving their consent, subjects indicated their age, gender, and state of residence and were screened in/out

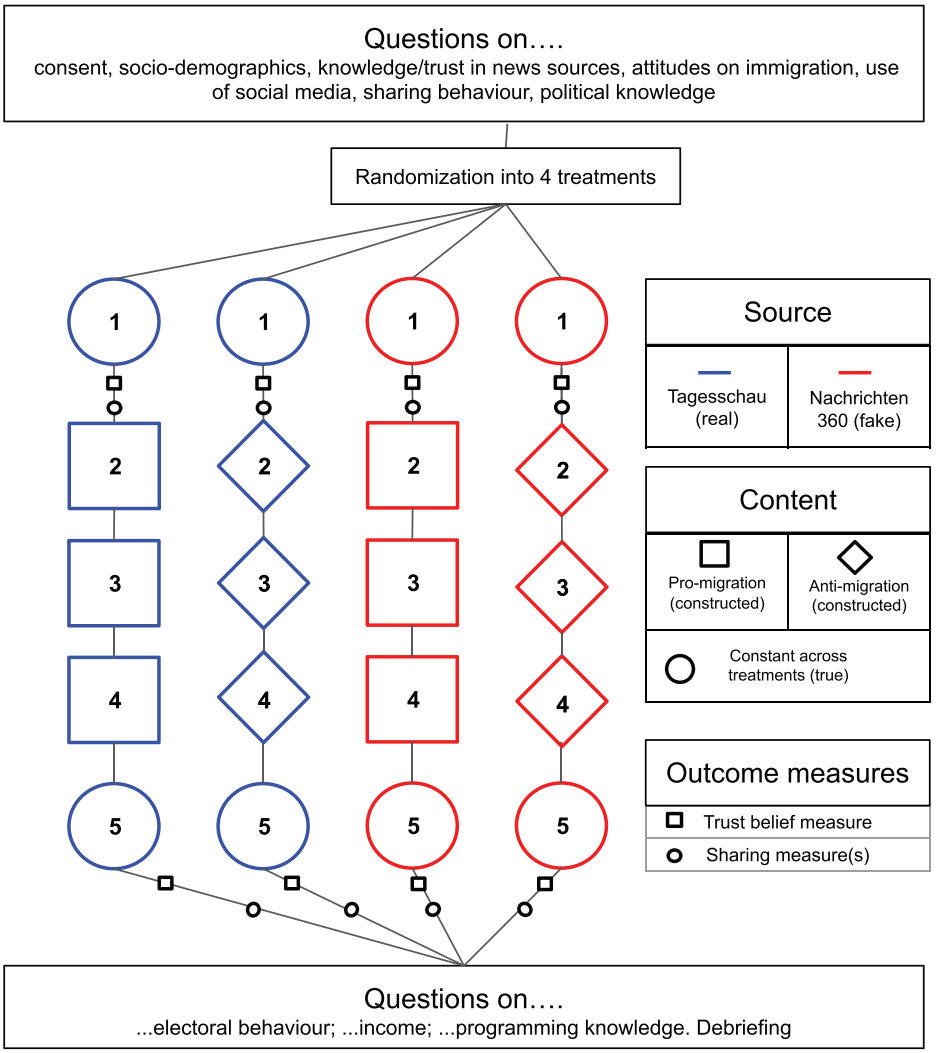

Figure 1. Experimental setup.

Participants read news reports 1–5; Report 1 and 5 are true and the same for all participants; Reports 2–4 are manipulated, either showing anti-immigration or pro-immigration content. Combined with respondents previous attitudes we construct a randomized treatment referring to either congruent or incongruent ideological content. The experiment contained a third randomized dimension which is not depicted here (see below and App. A.10.5)

accordingly. Subsequently, we measured a range of covariates (see Appendix Table A1 for the question wording): to measure familiarity with news sources, we exposed subjects to a list of eleven real or made-up outlet names in random order and asked whether they recognized the source and, if so, whether they had already read or watched news by that source. From the former question, we constructed a "mainstream source knowledge" index averaging scores for seven sources. This was followed by a standard trust question asked for the same list of sources. Again, we constructed a "mainstream media trust" index averaging scores for seven mainstream sources. Subsequently, we

measured attitudes on immigration with a five-item battery. The items showed high reliability (Cronbach's alpha = 0.93), and we computed an individual average across the five items.

We further asked respondents whether they had an account for e-mail, Facebook, Twitter and Whatsapp (see Figure A7) and how frequently they shared news via each (see Figure A9). Sharing daily or several times a week was most common via Whatsapp (28%), followed by Twitter (19.2%), Facebook (17.9%) and e-mail (14.9%). This question was followed by a battery measuring political knowledge and a question eliciting overconfidence in that knowledge.

As depicted in Figure 1, the ensuing experimental stage required subjects to read five news reports, Reports 1–5. The central task was to indicate their belief and sharing intention for the two true reports, i.e., Reports 1 and 5. All reports, whether true or manipulated, provided information about a specific factual question bearing on immigration and immigration policy, a topic likely to be salient to respondents.[3] Report 1 (true) presented current numbers of family reunions among refugees in Germany. Report 2 (manipulated) concerned representation of immigrants in crime statistics. Report 3 (manipulated) provided data on the performance of refugees in German language tests. Report 4 (manipulated) presented data on a possible pull effect of private sea rescue in the Mediterranean and Report 5 (true) concerned the use of emigration subsidies. To make the contrasts between sources meaningful, we had to assume that few subjects had read the true Reports 1 and 5 for which we measured our outcomes: as discussed previously, sources should matter when news receivers do not already know the truth. We therefore picked reports that had a low number of shares on social media, and for which we found little related reporting.[4]

To boost experimental realism, we presented reports exactly as they would look like online, i.e., as posts on the social media platform Facebook or articles on news websites, by providing a screenshot of the (allegedly) original report before the text. Figure 2 shows an example of such a screenshot and Appendix A.12 includes more examples.

We manipulated the reports along three dimensions. In the following, we refer exclusively to the two treatment dimensions relevant to our predictions. We discuss the third dimension, i.e., whether the news report was taken from Facebook or a news website, elsewhere. Corresponding to the two treatment dimensions discussed, participants were hence randomly assigned to four groups. Appendix A.4 provides balance statistics and shows that there are no significant differences across treatment conditions. Most importantly, immigration attitudes and voting intentions do not differ significantly across treatments.

Source Treatment

Across all five news reports, we varied the news source, randomly assigning participants to receive all five reports from either of two sources. We operationalized the *real* source using the name and appearance of the *Tagesschau*, the news section of the largest German public broadcaster, likely to be known by most people. For the fake source, we made up an outlet called *Nachrichten 360*. We chose a name that sounded neutral and without any partisan tendency, and made the logo and appearance look typical of real news organizations, thereby reflecting the logic of fake sources feigning credibility. As Figure A2 illustrates, the

Figure 2. Example of Facebook post screenshot accompanying reports.

The teaser translates as "The figures from the latest police statistics show: Immigrants are not suspected of a crime in Germany above average", the headline as "Immigrants not overrepresented among criminal suspects" and the beginning of the text as "Last year, 1.9% of all crime suspects … ".

real source is indeed known by most people. In contrast, few people imagine to know the fake source.

Content & Congruence Treatment

We further varied the content of the three middle reports, Report 2–4. Report 1 was a true (but slightly shortened) report and identical for all subjects. Thereafter, respondents were randomly assigned to read either a constructed "pro-immigration" or "anti-immigration" version of each of the Reports 2–4. To construct such two versions of a report, we departed from some piece of real data and manipulated it in two directions, so that one version would be convenient to subjects with an immigration-friendly outlook, and the other version to those with an immigration-skeptic outlook. For example, Report 2 provided the latest statistics on crime rates among immigrants as compared to natives. The *pro-immigration* treatment group read that rates were lower than in reality; the *anti-immigration* treatment group read that they were higher. Report 5, was again true (but slightly shortened) and the same for all respondents.

We combined the content randomization of Reports 2–4 with participants' immigration attitudes to construct a *congruence treatment*. The treatment is congruent if an immigration-skeptic respondent reads an *anti-immigration report* or if a supporter of immigration reads a *pro-immigration report*. The incongruent treatment consisted of the

analogous opposite cases. We defined those below the mean of the composite index as immigration skeptics, and those above the mean as immigration supporters. This choice is robust against an alternative implementation, defining immigration skeptics and supporters in terms of tertiles, as shown in Appendix A.10.1.

After Report 1 and 5, respectively, subjects were asked to indicate whether they believed the report was true (*On a scale from 0 to 6, do you think that the information in the text of [source] is true? 0 means not at all, 6 means completely*), and whether they would share the report with the question: *Would you share the message from [news source] that you just read via …*, followed by a small battery that included the options *E-Mail: Yes/No, Facebook: Yes/No Twitter: Yes/No Whatsapp: Yes/No*, depending on which services respondents had indicated using. Although this question measures hypothetical sharing decisions, we are confident that it is an approximation of real sharing behavior, as shown elsewhere (Mosleh et al., 2019). This measure of sharing intention has provided valuable insights in several recent studies (Pennycook et al., 2019; Pennycook & Rand, 2020).

Finally, we inquired about subjects' education, their turnout and vote choice at the previous election, income, and basic programming knowledge. Before participants could complete the survey, we debriefed and informed them about the purpose of the experiment and the corresponding manipulations. Our experimental design was approved by the Ethics Committee of the European University Institute under file number #CG8-1-2019. We employed several strategies to mitigate the impact of deception in our study: First, we clarified which of the sources and contents were constructed, and provided subjects with the true facts for the manipulated reports. Second, we emailed more substantive information related to the news reports to those interested a few weeks after the study. Third, we provided them with an open-ended feedback box after the debriefing, which revealed that the the survey experience was overwhelmingly positive. Please see Appendix A.11 for further discussion of ethical considerations.

Analysis

To test our preregistered hypotheses, we rely on t-tests for the main effects and a linear regression for the interaction effect. In addition, we conducted a whole series of robustness checks (cf. Section A.10) and manipulation checks to make sure participants received the treatment (cf. Section A.10.3), and we tracked their attention while answering the questionnaire with a JavaScript module developed by Diedenhofen and Musch (2017).

Regarding treatment heterogeneity discussed in Section 2.3, we did not prespecify any predictions. Because of the large number of covariates, testing for heterogeneity manually by either iteratively subsetting the data or including a large number of interaction terms would be cumbersome. More importantly, p-values would no longer be valid due to multiple testing issues (Athey & Imbens, 2017). We therefore follow an exploratory approach and rely on novel machine learning methods developed in Athey and Imbens (2016), Wager and Athey (2018), and Athey et al. (2019), which allows for testing heterogeneity across any number of covariates without running into validity issues. The "causal forest" approach is based on the fundamental concept of random forests as well as the generalized random forest framework proposed Athey et al. (2019). See Appendix A.13.1 for details.

Empirical Results

Source Treatment

Hypothesis H1a posits that *people have a higher tendency to believe news reports from a real source than from a fake source*. To test this, we compare belief in Report 1, which has the same content for all subjects, across source treatment groups. Plot A in Figure 3 visualizes group means and results from t-tests. The real source indeed elicited higher average belief (also see Table A11 for OLS estimates). In substantive terms, the difference of 0.62 on a 7-point scale is comparable to source effects in studies with similar designs (e.g., Knight Foundation/Gallup, 2018). It is, however, much smaller than individual-level predictors of belief such as familiarity with a story in Pennycook et al. (2018).

We also expected that people *would be more likely to share news reports that come from real sources than those that come from fake sources (H1b)*. Plot B in Figure 3 illustrates that intention to share is generally low. The plot also does not suggest large source effects on sharing. T-tests for e-mail (mean difference: 0.02) and Twitter (mean difference: 0.04) do not yield statistically significant effects. In the case of Twitter, any real difference would be difficult to detect due to the small number of Twitter users in our sample. In contrast, there are significant differences for Facebook (mean difference: 0.05) and Whatsapp (mean difference: 0.03).

Although we did not assume that the treatment effect on sharing must necessarily work through belief, we ran some additional analyses to explore this possibility. Figure A6 in the Appendix shows that belief is positively correlated to all sharing outcomes (about 0.2 for each correlation). If sharing could be fully explained by belief, we would expect a much higher correlation. The much lower correlation suggests that other factors are more important.

Effects on sharing seem small but could turn out substantial in the presence of cascade dynamics on social media (Del Vicario et al., 2016). The plots show a low level of sharing tendency in general. It is somewhat lower than in recent comparable studies in the US context, although these collapse several response categories (Pennycook et al., 2018, 2019; Pennycook et al, 2020).[5] However, once we look at results for "frequent sharers" only, we find that both baseline sharing and the source effect on Facebook sharing is larger (see Appendix A.10.2). We further looked at sharing intentions by partisanship. In contrast to observational studies in the US (e.g., Benkler et al., 2018), we do not find any clear patterns in partisan sharing behavior (see Appendix A.7).

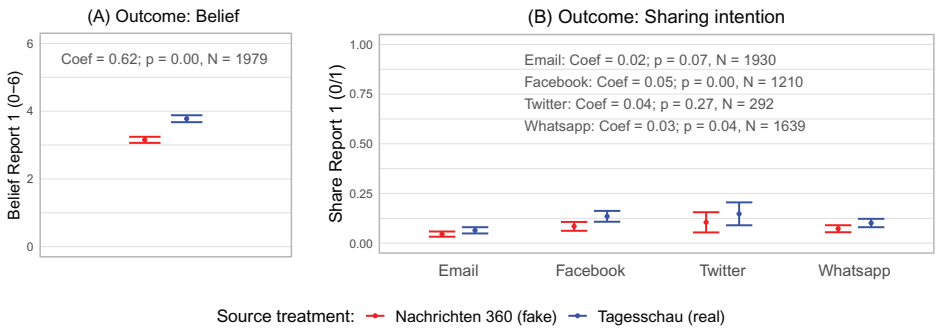

Figure 3. Source effect on news belief and sharing.

Congruence Effects

We further expected that *those who previously saw congruent (vs. incongruent) news reports by a source would have a higher tendency to believe a subsequent news report by the same source (H2a)* and that they would be *more likely to share the news report (H2b)*. We test this prediction for Report 5, which had the same content across treatments, after subjects had read three reports, either congruent or incongruent according to their pre-treatment attitudes.

Figure 4 visualizes average belief across the two treatments and results from t-tests (see Table A12 for OLS estimates). In support of H3a, subjects who saw three congruent reports indeed show a higher belief in the subsequent report (0.19 on a scale from 0 to 6). The effect is smaller than in studies that test the effect of congruence of information on believing the *same* information. For example, Kuru et al. (2017) find that a congruent report, compared to an incongruent report, is judged 0.14 units more true on average (on a scale from 0 to 1). However, in our comparison, the treatment difference is not due the content of Report 5, which is the same across treatments, but can only be explained by *previous* exposure to three news reports.

There is less evidence for effects of previous congruence on later sharing (H3b). T-tests for e-mail (mean difference: 0), Whatsapp (mean difference: 0.02), and Twitter (mean difference: 0.07) do not show statistically significant effects. Only for Facebook there seems to be a significant effect (mean difference: 0.05): this implies that people are 5% more likely to express a sharing intention of the same report when received by a source that has previously provided congruent content than by a source that has previously provided incongruent content. Again, we point the reader to correlations between belief and sharing of Report 5 as reported by Figure A6 in the Appendix. Positive but weak correlations imply that belief is not a strong factor for sharing intentions.

Source–congruence Interaction Effects

We further hypothesized that *difference in later news belief between those who saw congruent and those who saw incongruent reports would be higher for the fake source than for the real source (H3a)* and expected the same for *sharing intentions (H3b)*. Figure 5 visualizes the means in the four treatment groups as well as OLS estimates for an interaction effect (see also Table A13). For belief, the significant interaction term

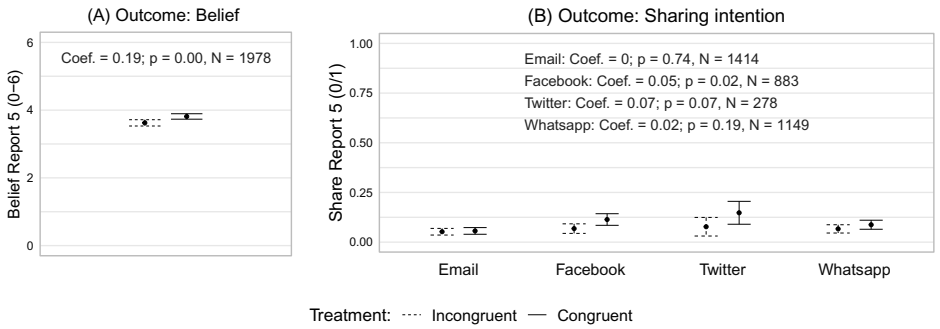

Figure 4. Congruence treatment, belief and sharing.

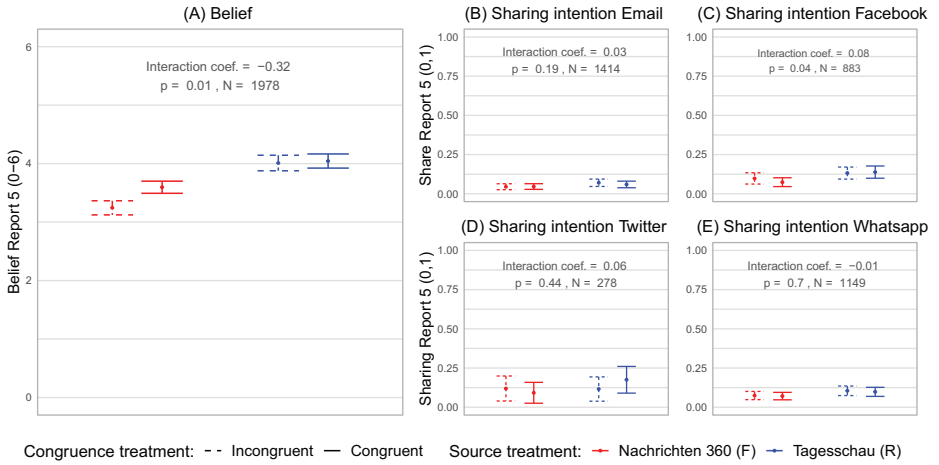

Figure 5. Interactions between source and congruence treatment.

(b = −0.32) supports our prediction. The positive regression coefficients for congruence and source and the negative interaction coefficient imply that, for the fake source (source treatment variable = 0), the effect of congruence on belief is 0.35. In contrast, the effect of congruence is negligible for the real source (0.03). In substantive terms, the effect of congruence for the fake source (0.35) is more than half of the main source effect noted above (0.62). In other words, the fake source can "make up" about half of the gap in belief after only three congruent stories. However, we do not know if and at what point this trend would lead to convergence. For our sharing outcomes, there is no clear pattern, with Facebook showing a small but significant interaction in the opposite direction.

Since we measured belief of both Report 1 and Report 5, we can interpret the results in terms of changing credibility as illustrated in Figure 6. The difference in belief of Report 1 between the two sources (*Nachrichten 360* and *Tagesschau*) was due to higher credibility of the real source (*Tagesschau*). Subsequently, for the real source (*Tagesschau*) it does not matter much whether readers receive congruent or incongruent reports, as the non-difference for Report 5 shows that credibility perceptions did not change for the real source. In contrast, respondents had a higher tendency to believe Report 5 by the fake source (*Nachrichten 360)* if it had provided them with congruent news. The fake source thus gained in credibility.

Note that Figure 6 gives the impression that belief increases *overall* between Reports 1 and 5. One reason could be that Report 1 and Report 5 are not equally plausible. Perhaps, Report 5 was just easier to square with people's preexisting political knowledge than Report 1. It could also be that it was a "mere exposure" effect: subjects get used to the source and trust it more, whatever it does. In any case, the difference for the fake source can only be attributed to the randomly assigned exposure to previous reports.[6]

Treatment Heterogeneity

We now turn to the question of treatment heterogeneity for the two treatment dimensions, i.e., source and congruence. We estimate heterogeneity using the causal forest

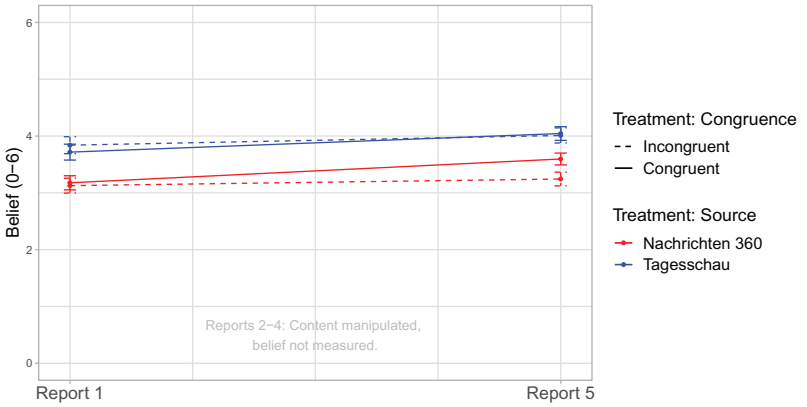

Figure 6. Source credibility development.

method (cf. Wager & Athey, 2018). An elaborate description of this approach can be found in Appendix A.13.1. Results for the belief outcome are described in this section; results for sharing outcomes can be found in Appendix A.13.3.

Following Athey and Wager (2019), we generally start by growing a pilot causal forest based on all covariates described in our methods section. The variables included are gender, age, state of residence (coded as dummy for Eastern Germany), mainstream media knowledge, mainstream media trust, political knowledge, overconfidence, use of e-mail/Facebook/Twitter/Whatsapp (coded as dummies), sharing frequency, education, turnout at last election, vote choice at last election (coded as party dummies), income, and basic programming knowledge. We then regrow the causal forest including only the variables with above-average importance in the pilot forest (cf. Athey & Wager, 2019).[7] For this final forest, we predict individual-level treatment effects, i.e., the difference between being exposed to one treatment compared to the other, with out-of-bag prediction.

The resulting causal forest for the source treatment is built with N = 1751 observations, on eight covariates. Before discussing the significance of each covariate, we report the results of two omnibus tests whether heterogeneity is generally present or not. Again following Athey and Wager (2019), we first compare regions with high and low estimates of individual treatment effects, separated by the median. In the absence of heterogeneity, we should not find large differences across these regions. In our case, there is a significant difference of 0.53 (SE = 0.13) between the averages of these two region. A second omnibus test relies on the best linear prediction method that fits the the conditional average treatment effect as a linear function of the the causal forest estimates. The significant coefficients (p = .000 suggest that heterogeneity is present, see Appendix A.13 for details).

Figure 7 visualizes heterogeneity by plotting predictions against the set of eight covariates. Recall that the average treatment effect of the source treatment on belief was 0.62. This effect varies considerably with a few covariates. For example, Panel A shows that, for subjects expressing high trust in mainstream media, the predicted difference between real and fake source is about 0.75 points higher than for those with low mainstream media trust. Panel B shows that the real source made a greater impression on those

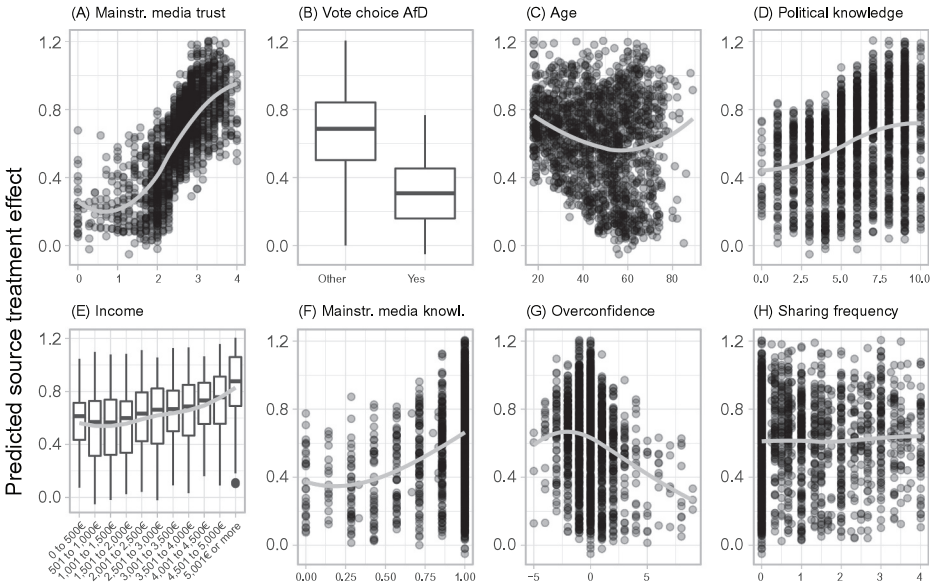

Figure 7. Heterogeneity of predicted source treatment effect on belief.

who did not vote for the right-wing populist party AfD. The remaining plots suggest that the source effect is weakest for middle-aged people (C), for those least politically knowledgeable (D), for those who know few mainstream sources (E), for those with lower incomes (F), and for those overly confident of their political knowledge (G). The treatment effect does not seem to vary much with individual sharing frequency. To test whether these relations are significant, we run a best linear projection of predictions on the set of covariates (cf. Tibshirani et al., 2020). It shows that only three variables are statistically significant at the 95% level: Mainstream media trust (p = .01), vote for the populist right (p = .005), and political knowledge (p = .035). Age and media knowledge show marginal significance levels (see Table A27 in the Appendix for more details).

To test treatment heterogeneity for the congruence effect, we restricted the data to subjects in the fake source treatment, as it is here that we found a congruence effect. We grew a causal forest as described above, based on N = 888 observations and, again, eight covariates with above-average importance. The set of most important covariates is similar to those important for the source effect. The two omnibus tests described above suggest that the heterogeneity found is real: the average difference between the high-prediction region and the low-prediction region is 0.397 (SE = 0.166). The calibration test reveals a significant coefficient (p = .047).

Figure 8 visualizes heterogeneity of the congruence effect against the included covariates. Recall that the average treatment effect was 0.35. For example, Plot (B) suggests that the congruence effect is larger for middle-aged people. The other plots suggest substantial heterogeneity for vote for the populist right, income, political knowledge, and education. Again, we use the best linear projection method to test whether covariates significantly

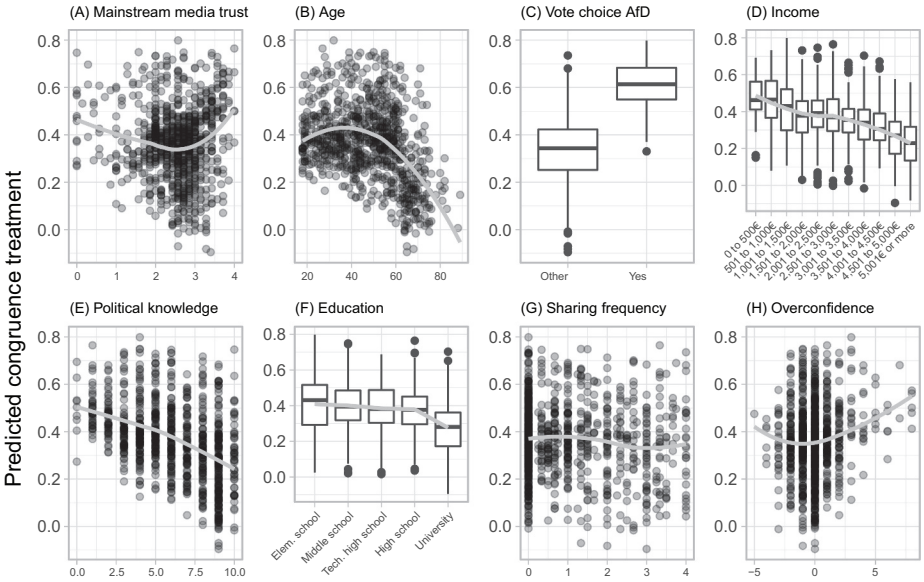

Figure 8. Heterogeneity of predicted congruence treatment effect on belief.

predict heterogeneity. Only having voted for the populist right turns out to be statistically significant at the 95%-level (p = .001, see Table A27 in the Appendix for more details).

These results imply some interesting patterns: by and large, there is an inverse relation between the heterogeneity of the source effect and of the congruence effect. For example, those not having voted for the right-wing populist AfD give comparably greater credit to the real source than to the fake source, but they are less influenced by a source providing them congruent information. A similar conclusion could be drawn about mainstream source trust, political knowledge, and income, although we have less confidence about congruence effect heterogeneity on these covariates.

We also explored the heterogeneity of both treatments for the sharing outcomes. However, both omnibus tests do not allow us to reject the hypothesis of no heterogeneity, so that we do not further explore the significance of individual covariates. This somewhat reflects the much less clear-cut findings for sharing in the main analyses and could potentially also be due to the fact that our samples become small in the high-dimensional space defined by our covariates, as discussed below. See Appendix A.13.3 for more details.

Discussion and Conclusion

With the advent of the Internet and social media, the number of sources for political information has multiplied. A particular worrying aspect is that malevolent actors can – and do – easily invent sources that present themselves as legitimate online news organizations. In a novel survey experiment, we investigate how source characteristics and content affect people's belief that the factual claims contained in a news report are true and their intention to share the report. In line with our first hypothesis, we find that individuals

have a stronger belief in the report by a real than by a fake source, as well as a higher intention to share it, at least via Facebook and Whatsapp. However, the gap between the real source and the fake source is not particularly large. Additional evidence shown in the Appendix (see Figure A3) suggests that the naming alone of our made-up source made a difference. A worrying implication of our finding is that people might be easily fooled by malevolent actors who get such tiny signals right.

Our key finding concerns the role of attitudinal congruence of content. Being exposed to a series of reports that present facts congruent (rather than incongruent) with one's attitudes increases belief in and Facebook sharing of a subsequent report of that same source. Interestingly, we find this congruence effect exclusively for the fake source. Hence, while belief in a real source with an arguably widely known reputation seems less malleable, fake sources may manage to build up credibility through catering to readers' world views. Given that this effect could be seen after only three news reports in our experiment, the potential for fake news entrepreneurs or hyperpartisan sources to win people's trust might be substantial – whether through selective reporting or outright misinformation.

Furthermore, we show that the effects on belief vary across subpopulations. The positive effect of obtaining information from our real source on belief seems to be moderated by distrust of the mainstream media, vote intention for the populist right, and political knowledge. This parallels other recent evidence that people low on media trust are more susceptible to disinformation (Zimmermann & Kohring, 2020). Similarly, the effect of congruence varies in particular across vote intention for the populist right. Considering the anti-elitist dimension of populism (e.g., Schulz et al., 2018), it is not surprising that voters of the populist right seem to quickly develop trust in a "new" and ideologically convenient source.

Several details and limitations of our experiment warrant discussion and open avenues for future research. First, our findings suggest belief and sharing are distinct outcomes. As far as the relationship between the two variables is concerned, we find relatively low correlations between belief and sharing intention (see Figure A6). We suggest that sharing is explicable only to a small extent by belief, as implied by other recent studies (Bright, 2016; Pennycook et al., 2019). To what extent sharing is caused by belief should be answered by designs that, in contrast to our study, causally identify belief as an independent variable, and separate it from other factors (e.g., Cappella et al., 2015; Valenzuela et al., 2017). Importantly, sharing decisions are affected by various factors that may be less important for belief, such as an item's newsworthiness, informational utility, valence, and framing. For example, recent studies have shown that content with moral frames and language are more likely to be shared (Valenzuela et al., 2017). This could be one reason why our results for belief and sharing intentions differ. In our study, we only create factual differences between treatment stimuli. Future studies should combine manipulations of facts with manipulations of other content-level variables such as the framing of information.

The lack of effects on sharing might also be due to statistical power. Our sample was recruited according to nationally representative quotas. This inevitably meant including many subjects who do not use some of the social media services. As we only measured sharing outcomes for users of the respective platforms, sample sizes for the analyses of sharing intention were much smaller. This was especially the case for Twitter. However, we note that the findings for Facebook seem robust despite a smaller sample size. Future

work should collect larger samples of social media users (or oversample them) to have enough power to identify small effects.

Second, our design was not focused on identifying the psychological mechanisms, i.e., the thought processes underlying belief or sharing. Rather, we approached the concept of source credibility from the perspective of the source. Given that anyone can quickly set up what appears to be a news source through a web site or a Facebook page, we asked to what extent people make a difference between a real source and a fake source imitating a credible outlet. This approach meant that we did not delve into perceptions about the sources. For example, it could be that subjects thought the made-up source was a legitimate news organization, but unknown to them; alternatively, they may have assumed it was a site potentially peddling misinformation. Although we ultimately can not know how participants perceived the sources at the experimental stage, our data suggests that subjects in our experiment judged the made-up source as unfamiliar but to some extent credible. Pretreatment measures of familiarity and trust over a range of sources suggest that even though most subjects rightly said they did not know the made-up source, they still rated it as relatively trustworthy (e.g., better than the tabloid *Bild*; see Figures A2 and A3). Slight differences in the name of a source may have strong effects, as implied by the difference in trust in the made-up sources *Nachrichten 360* (used in our experiment) and *Berliner Nachrichten* (cf. Figure A3 in the Appendix). To gain a better understanding of the psychological underpinnings that cause such differences, future studies should further explore the experimental variation of sources' names, appearance, and content in a combination with open-ended probing questions. Eye tracking could provide a fruitful way to measure to what extent individuals perceive and discern different signals (Sülflow et al., 2019).

Third, our choice of the topic – immigration – warrants discussion. While we are confident that there was nothing extraordinarily controversial about the topic at the particular time of the study (see our explorations of Google Trends in Appendix A.14), immigration has been a relatively salient and controversial topic in recent years. We cannot say with certainty that our findings – especially that a source gains credibility when it provides congruent information – would show for other topics. However, we came to our theoretical prediction on the basis of motivated reasoning theory. If people have a tendency to defend their preexisting attitudes, a source helping them to do so will gain credibility in their view. Many scholars have argued that motivated reasoning should occur in particular for issues of ideological conflict (e.g., Kahan, 2016a; Slothuus & de Vreese, 2010; Taber et al., 2009). This seems to suggest that our findings apply to other controversial issues. Nonetheless, future work should test the scope of our findings.

Finally, in our experiment we neglected individuals' choices of both news sources and content. The interplay of source selection with media trust (Fletcher & Park, 2017; Tsfati & Cappella, 2005) as well as with extreme attitudes (Bakshy et al., 2015; Eady et al., 2019; Flaxman et al., 2016; Gentzkow & Shapiro, 2011) is complicated. Individual choice on social media platforms also interacts with a platform's technology. Accordingly, whether the effects studied are reduced or even magnified by individual choice is a central question for future research.

Our findings have implications for the spread of misinformation on social media. Platforms play a key role in directing users to fake-news websites (Guess, Nyhan et al.,

2020b). Relying on the mechanism we identified, such websites may try to exploit recommendation algorithms (that show users what they like) to build a "relationship" of trust with their readers and viewers. Therefore, once a source has been identified as questionable by a platform, further contact between source and user should be cut off. The source should also be prevented from initiating contact with further users. Once a source–user relationship continues outsides of the platform, it is too late for platforms to react. Our findings also imply that small signals such as a source's name have a strong impact. We motivated our study with the example of the "Denver Guardian", a scam outlet ostensibly playing on existing trusted news brands and familiar places. Platforms' filtering algorithms should be adapted to detect the use of such tiny signals as well as more abstract strategies underlying them. Our results support recent attempts by platforms to integrate source- and network-level characteristics into automated detection mechanisms (Gereme & Zhu, 2019). In addition, platforms should educate users as to how sources may fool them into believing and sharing misinformation. Making users aware of the psychological mechanisms that increase their vulnerability, possibly through playful demonstrations of how to spot misinformation, could help (Guess et al. 2020a). In a battle between automated detection tools and misinformation entrepreneurs that tailor their strategies, platforms have to develop solutions that are adaptive and evolve over time.

Acknowledgments

We are grateful to the participants of several workshops and colloquiums at the European University Institute (May 2019), ISPP Lisbon (July 2019), Massachusetts Institute of Technology (MIT) (October 2019) and New York University (December 2019) for feedback. We are especially grateful to David Rand, Gordon Pennycook, Joshua Tucker, John Jost, Johanna Gereke, Wiebke Drews, Simon Munzert, Max Schaub, Anselm Hager, Tobias Widmann, Christian Clemm, Florian Keusch, Matthias Fatke, Kathrin Ackermann, Christine Emmer, Nan Zhang and several anonymous reviewers for excellent feedback. Code and data to reproduce the study are available here: https://doi.org/10.7910/DVN/EJYWCK.

Notes

1. There is some confusion about how the concepts of "media trust" and "source credibility" relate (Kohring & Matthes, 2007). In our terminology, both "credibility" and "trustworthiness" describe a news receiver's perception of the source. When someone perceives a source as credible or trustworthy, she "trusts" that source. The concept of "media trust" describes average credibility/trustworthiness perceptions across a set of sources; "mainstream media trust", the average perception across sources considered mainstream.
2. There is a debate whether the effect of congruence can be explained within a Bayesian framework or whether it should be conceived of as "bias", which is not crucial for our expectations below (Gerber & Green, 1999; Kahan, 2016b; Tappin et al., 2018).
3. One possible concern is that we conducted our experiment during times in which those issues were particularly controversial. Google trends data seems to suggest otherwise (see Appendix A.14, Figure A19 and A20). Regarding the issue of immigration, while their concrete policy positions vary, Germany's parties can be broadly categorized in immigration-critical parties (AfD, CDU/CSU) and parties that are more supportive of immigration (SPD, Greens, Left Party).

4. Note that we opted to measure the outcomes on (slightly modified) true rather than fake news reports. It would also be interesting to understand how our treatments affect believing and sharing of false content. However, the plausibility of the chosen fake news report is hard to calibrate. Had we inadvertently picked a very implausible piece of misinformation, people would not believe it irrespective of the source. This is less problematic for a true news report, because it represents an aspect of the real world.
5. Sharing intentions reported by the cited studies include both subject answering "yes" and subject answering "maybe".
6. Note further that Figure 6 also shows slight differences in belief between the congruence and incongruence treatment groups *before* the treatment is delivered. To assess the size of these initial differences, we ran t-tests within the source treatment groups comparing belief in Report 1 of subjects who received the congruent and incongruent treatments. For both the fake ($p = 0.586$) and the real source ($p = 0.241$), the initial differences at Report 1 are likely due to chance. We also reran the main analyses using the subsample of respondents who passed all our manipulation checks and found that those initial differences at Report 1 are much smaller once we take this subsample as visualized in Figure A11.
7. Importance is calculated as a simple weighted sum of how many times a variable was split on at each depth in the forest. Importance statistics for all variables can be found in Appendix A.13.

Notes on contributors

Paul C. Bauer is a research fellow at the Mannheim Centre for European Social Research and pursues research in the areas of political behavior, political communication, and computational social science. His work has appeared in Public Opinion Quarterly, Political Behavior, PLOS One, and the European Sociological Review among others.

Bernhard Clemm von Hohenberg is a PhD candidate at the European University Institute. His research extends across political communication, political psychology, computational social science, and experimental methodology.

ORCID

Paul C. Bauer http://orcid.org/0000-0002-8382-9724
Bernhard Clemm von Hohenberg http://orcid.org/0000-0002-6976-9745

References

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, *31*(2), 211–236. https://doi.org/10.1257/jep.31.2.211

American Press Institute. (2017). *'My' media versus 'the' media: Trust in news depends on which news media you mean*. https://www.americanpressinstitute.org/publications/reports/survey-research/my-media-vs-the-media/

Arceneaux, K., Johnson, M., & Murphy, C. (2012). Polarized political communication, oppositional media hostility, and selective exposure. *Journal of Politics*, *74*(1), 174–186. https://doi.org/10.1017/S002238161100123X

Athey, S., & Imbens, G. (2016). Recursive partitioning for heterogeneous causal effects. *PNAS*, *113*(27), 7353–7360. https://doi.org/10.1073/pnas.1510489113

Athey, S., & Imbens, G. W. (2017). The econometrics of randomized experiments. In Esther Duflo, Abhijit Banerjee *Handbook of economic field experiments* (Vol. 1, pp. 73–140). Elsevier.

Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, *47*(2), 1148–1178. https://doi.org/10.1214/18-AOS1709

Athey, S., & Wager, S. (2019). *Estimating treatment effects with causal forests: An application*. http://arxiv.org/abs/1902.07409

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. *Science*, *348*(6239), 1130–1132. https://doi.org/10.1126/science.aaa1160

Barberá, P., Wang, N., Bonneau, R., Jost, J. T., Nagler, J., Tucker, J. A., & González-Bailón, S. (2015). The critical periphery in the growth of social protests. *PLoS One*, *10*(11), e0143611. https://doi.org/10.1371/journal.pone.0143611

Baum, M. A., & Groeling, T. (2009). Partisan cues and public opinion regarding national security and war. *Political Behavior*, *31*(2), 157–186. https://doi.org/10.1007/s11109-008-9074-9

Baum, M. A., & Gussin, P. (2004). Quarterly Journal of Political Science. *In the eye of the beholder: How information shortcuts shape individual perceptions of bias in the media*, 3(1), 1-31. http://dx.doi.org/10.1561/100.00007010

Benkler, Y., Faris, R., & Roberts, H. (2018). *Network propaganda: Manipulation, disinformation, and radicalization in American politics*. Oxford University Press.

Berelson, B. R., Lazarsfeld, P. F., & McPhee, W. N. (1954). *Voting: A study of opinion formation in a presidential campaign*. University of Chicago Press.

Berghel, H. (2017). Lies, damn lies, and fake news. *Computer*, *50*(2), 80–85. https://doi.org/10.1109/MC.2017.56

Boczkowski, P. J., & Mitchelstein, E. (2012). How users take advantage of different forms of interactivity on online news sites: Clicking, E-mailing, and commenting. *Human Communication Research*, *38*(1), 1–22. https://doi.org/10.1111/j.1468-2958.2011.01418.x

Bright, J. (2016). The social news gap: How news reading and news sharing diverge. *Journal of Communication*, *66*(3), 343–365. https://doi.org/10.1111/jcom.12232

Cappella, J. N., Kim, H. S., & Albarracín, D. (2015). Selection and transmission processes for information in the emerging media environment: Psychological motives and message characteristics. *Media Psychology*, *18*(3), 396–424. https://doi.org/10.1080/15213269.2014.941112

Chadwick, A., & Vaccari, C. (2019). *News sharing on UK social media: Misinformation, disinformation, and correction*. Loughborough: Online Civic Culture Centre, Loughborough University.

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*(5), 752–766. https://doi.org/10.1037/0022-3514.39.5.752

Chebat, J.-C., Filiatrault, P., Laroche, M., & Watson, C. (1988). Compensatory effects of cognitive characteristics of the source, the message, and the receiver upon attitude change. *Journal of Psychology*, *122*(6), 609–621. https://doi.org/10.1080/00223980.1988.9915535

CHI '20: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, April 2020, Pages 1–11. https://doi.org/10.1145/3313831.3376232

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). The spreading of misinsformation online. *Proceedings of the National Academy of Sciences*, *113*(3), 554–559. https://doi.org/10.1073/pnas.1517441113

Delli Carpini, M. X., & Keeter, S. (1996). *What Americans know about politics and why it matters*. Yale University Press.

Diedenhofen, B., & Musch, J. (2017). Pagefocus: Using paradata to detect and prevent cheating on online achievement tests. *Behavior Research Methods*, *49*(4), 1444–1459. https://doi.org/10.3758/s13428-016-0800-7

Ditto, P. H., Liu, B. S., Clark, C. J., Wojcik, S. P., Chen, E. E., Grady, R. H., Celniker, J. B., & Zinger, J. F. (2018). At least bias is bipartisan: A meta-analytic comparison of partisan bias in liberals and conservatives. *Perspectives on Psychological Science*.

Druckman, J. N. (2001). Using credible advice to overcome framing effects. *Journal of Law, Economics, and Organization*, *17*(1), 62–82. https://doi.org/10.1093/jleo/17.1.62

Druckman, J. N. (2014). Pathologies of studying public opinion, political communication, and democratic responsiveness. *Political Communication*, *31*(3), 467–492. https://doi.org/10.1080/10584609.2013.852643

Eady, G., Nagler, J., Guess, A., Zilinsky, J., & Tucker, J. A. (2019). How many people live in political bubbles on social media? Evidence from linked survey and Twitter data. *SAGE Open*, *9*(1), 2158244019832705. https://doi.org/10.1177/2158244019832705

Faz. (2020). *Sorge Vor Epidemie: Coronavirus-Krisenstab Sucht 300 Karnevalsbesucher*. Frankfurter Allgemeine Zeitung GmbH. https://www.faz.net/1.6652280

Flanagin, A. J., & Metzger, M. J. (2000). Perceptions of internet information credibility. *Journalism & Mass Communication Quarterly*, *77*(3), 515–540. https://doi.org/10.1177/107769900007700304

Flaxman, S. R., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, *80*(Special Issue), 298–320. https://doi.org/10.1093/poq/nfw006

Fletcher, R., & Park, S. (2017). The impact of trust in the news media on online news consumption and participation. *Digital Journalism*, *5*(10), 1281–1299. https://doi.org/10.1080/21670811.2017.1279979

Fogg, B. J., Marshall, J., Laraki, O., Osipovich, A., Varma, C., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P., Treinen, M., (2001). What makes web sites credible? A report on a large quantitative study chi 2001. *Proceedings of the Sigchi Conference on Human Factors in Computing Systems*, *3*, 61–68.1. https://doi.org/10.1145/365024.365037

Gallup. (2019). *Gallup Poll social series: Governance*.Gallup News Service. https://int.nyt.com/data/documenthelper/6457-gallup-poll-social-series-gov/2647bae2b32ea42cdbe0/optimized/full.pdf

Gentzkow, M., & Shapiro, J. M. (2006). Media bias and reputation. *Journal of Political Economy*, *114*(2), 280–316. https://doi.org/10.1086/499414

Gentzkow, M., & Shapiro, J. M. (2011). Ideological segregation online and offline. *Quarterly Journal of Economics*, *126*(4), 1799–1839. http://arxiv.org/abs/arXiv:1011.1669v3

Gerber, A., & Green, D. (1999). Misperceptions about perceptual bias. *Annual Review of Political Science*, *2*(1), 189–210. https://doi.org/10.1146/annurev.polisci.2.1.189

Gereme, F. B., & Zhu, W. (2019). *Early detection of fake news 'before it flies high'*. Proceedings of the 2nd International Conference on Big Data Technologies (pp. 142–148). ICBDT2019. New York, NY: Association for Computing Machinery.

Goldman, S. K., & Mutz, D. C. (2011). The friendly media phenomenon: A cross- national analysis of cross-cutting exposure. *Political Communication*, *28*(1), 42–66. https://doi.org/10.1080/10584609.2010.544280

Greer, J. D. (2003). Evaluating the credibility of online information: A test of source and advertising influence. *Mass Communication and Society*, *6*(1), 11–28. https://doi.org/10.1207/S15327825MCS0601_3

Guess, A., Nyhan, B., & Reifler, J. (2018). Selective exposure to misinformation: Evidence from the. *European Research Council*, (pp. 1–14). European Research Council. https://www.semanticscholar.org/paper/Selective-exposure-to-misinformation%3A-Evidence-from-Guess-Nyhan/a795b451b3d38ca1d22a6075dbb0be4fc94b4000

Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020a, June). *A digital media literacy intervention increases discernment between mainstream and false news in the United States and India*. Proc. Natl. Acad. Sci. U. S. A.

Guess, A. M., Nyhan, B., & Reifler, J. (2020b). Exposure to untrustworthy websites in the 2016 US election. *Nature Human Behaviour*, *4*(5), 472–480. https://doi.org/10.1038/s41562-020-0833-x

Gunther, A. C. (1992). Biased press or biased public? Attitudes toward media coverage of social groups. *Public Opinion Quarterly*, *56*(2), 147. https://doi.org/10.1086/269308

Hargittai, E., & Walejko, G. (2008). The participation divide: Content creation and sharing in the digital age. *Information, Communication & Society*, *11*(2), 239–256. https://doi.org/10.1080/13691180801946150

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion*. Yale University Press.

Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *The Public Opinion Quarterly*, *15*(4), 635–650. https://doi.org/10.1086/266350

Iyengar, S., & Hahn, K. S. (2009). Red media, blue media: Evidence of ideological selectivity in media use. *Journal of Communication*, *59*(1), 19–39. https://doi.org/10.1111/j.1460-2466.2008.01402.x

Jerit, J., & Barabas, J. (2012). Partisan perceptual bias and the information environment. *The Journal of Politics*, *74*(3), 672–684. https://doi.org/10.1017/S0022381612000187

Jost, J. T. (2017). Ideological asymmetries and the essence of political psychology. *Political Psychology*, *38*(2), 416–417. https://doi.org/10.1111/pops.12407

Kahan, D. M. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, *8*(4), 407–424. http://journal.sjdm.org/13/13313/jdm13313.pdf

Kahan, D. M. (2016a). The politically motivated reasoning paradigm, part 1: What politically motivated reasoning is and how to measure it. In R. Scott & S. Kosslyn (Eds.), *Emerging trends in social and behavioral sciences* (pp. 1–6). John Wiley & Sons. https://doi.org/10.1002/9781118900772.etrds0417

Kahan, D. M. (2016b). The politically motivated reasoning paradigm, part 2: Unanswered questions. In R. Scott & S. Kosslyn (Eds.), *Emerging trends in the social and behavioral sciences* (pp. 1–15). John Wiley & Sons, Inc.

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, *1*(1), 54–86. Yale law school, public law working paper no. 307. https://doi.org/10.1017/bpp.2016.2

Kiousis, S. (2001). Public trust or mistrust? Perceptions of media credibility in the information age. *Mass Communication and Society*, *4*(4), 381–403. https://doi.org/10.1207/S15327825MCS0404_4

Knight Foundation/Gallup. (2018). *Perceived accuracy and bias in the media*.

Koehler, J. J. (1993). The influence of prior beliefs on scientific judgements of evidence quality. *Organizational Behavior and Human Decision Processes*, *56*(1), 28–55. https://doi.org/10.1006/obhd.1993.1044

Kohring, M., & Matthes, J. (2007). Trust in news media. Development and validation of a multidimensional scale. *Communication Quarterly*, *34*(2), 231–252. https://doi.org/10.1177/0093650206298071

Kuklinski, J. H., & Hurley, N. L. (1994). On hearing and interpreting political messages: A cautionary tale of citizen cue-taking. *The Journal of Politics*, *56*(3), 729–751. https://doi.org/10.2307/2132190

Kuru, O., Pasek, J., & Traugott, M. W. (2017). Motivated reasoning in the perceived credibility of public opinion polls. *Public Opinion Quarterly*, *81*(2), 422–446. https://doi.org/10.1093/poq/nfx018

Lippmann, W. (1922). *Public opinion*. Martino Publishing.

Loges, W. E., & Jung, J.-Y. (2001). Exploring the digital divide: Internet connectedness and age. *Communication Research*, *28*(4), 536–562. https://doi.org/10.1177/009365001028004007

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*(11), 2089–2109. https://doi.org/10.1037/0022-3514.37.11.2098

MacCoun, R., & Paletz, S. (2009). Citizens' perceptions of ideological bias in policy research. *Political Psychology*, *30*(1), 43–65. https://doi.org/10.1111/j.1467-9221.2008.00680.x

Metzger, M. J., Flanagin, A. J., Eyal, K., Lemus, D. R., & McCann, R. (2003). Credibility in the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. In P. Kalbfleisch (Ed.), *Communication yearbook 27* (pp. 293–335). Lawrence Erlbaum.

Mills, J., & Aronson, E. (1965). Opinion change as a function of the communicator's attractiveness and desire to influence. *Journal of Personality and Social Psychology*, *1*(2), 173–177. https://doi.org/10.1037/h0021646

Mondak, J. J. (1993). Source cues and policy approval: The cognitive dynamics of public support for the reagan Agenda. *American Journal of Political Science*, *37*(1), 186–212. https://doi.org/10.2307/2111529

Mosleh, M., Pennycook, G., & Rand, D. (2019). Self-reported willingness to share political news articles in online surveys correlates with actual sharing on Twitter. *PsyArXiv Working Paper*, 1–9.

Oeldorf-Hirsch, A., & DeVoss, C. L. (2019). Who posted that story? Processing layered sources in Facebook news posts. *Journalism & Mass Communication Quarterly, 97*(1), 141-160. https://doi.org/10.1177/1077699019857673

Page, B. I., Shapiro, R. Y., & Dempsey, G. R. (1987). What moves public opinion? *The American Political Science Review*, *81*(1), 23–44.https://doi.org/10.2307/1960777

Pennycook, G., Bear, A., Collins, E. & Rand, D G.. (2020). The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Headlines Increases Perceived Accuracy of Headlines Without Warnings. Management Science, 1–14. https://doi.org/10.1287/mnsc.2020.3478 11 66 doi:10.1287/mnsc.2019.3478

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology. General*, *147*(12), 1865–1880. https://doi.org/10.1037/xge0000465

Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2019). *Understanding and reducing the spread of misinformation online*. 10.31234/osf.io/3n9u8

Pennycook, G., & Rand, D. G. (2019a). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*, *88*, 185–200.10.1111/jopy.12476

Pennycook, G., & Rand, D. G. (2019b). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188* (September 2017), 39–50. https://doi.org/10.1016/j.cognition.2018.06.011

Pennycook, G., & Rand, D. G. (2019c). Fighting misinformation on social media using crowdsourced judgments of news source quality. Proceedings of the National Academy of Sciences, 116, 2521–2526. https://doi.org/10.1073/pnas.1806781116

Pennycook, G., Bear, A., Collins, E. & Rand, D. G. (2020). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science*, 1–14. Articles i. https://doi.org/10.1287/mnsc.2019.3478

Petersen, M. B., Osmundsen, M., & Arceneaux, K. (2018). *A need for Chaos and the Sharing of Hostile Political Rumors in Advanced Democracies*. DOI 10.31234/osf.io/6m4ts

Petty, R. E., Cacioppo, J. T., & Goldman, R. (1981). Personal involvement as a determinant of argument-based persuasion. *Journal of Personality and Social Psychology*, *41*(5), 847–855. https://doi.org/10.1037/0022-3514.41.5.847

Popkin, S. L. (1991). *The reasoning voter: Communication and persuasion in presidential campaigns*. University of Chicago Press.

Pornpitakpan, C. (2004). The persuasiveness of source credibility: A critical review of five decades' evidence. *Journal of Applied Social Psychology*, *34*(2), 243–281. https://doi.org/10.1111/j.1559-1816.2004.tb02547.x

PricewaterhouseCoopers. (2018). *Vertrauen in Medien 2018*.

Schulz, A., Mu, P., Schemer, C., Wirz, D. S., Wettstein, M., & Wirth, W. (2018). Measuring populist attitudes on three dimensions. *International Journal of Public Opinion Research*, *30*(2), 316–326. https://doi.org/10.1093/ijpor/edw037

Seidman, G. (2013). Self-presentation and belonging on Facebook: How personality influences social media use and motivations. *Personality and Individual Differences*, *54*(3), 402–407. https://doi.org/10.1016/j.paid.2012.10.009

Silverman, C. A. (2016). *This Analysis Shows How Viral Fake Election News Stories Out- performed Real News on Facebook*. https://www.buzzfeednews.com/article/craigsilverman/viral-fake-election-news-outperformed-real-news-on-facebook

Slothuus, R., & de Vreese, C. H. (2010). Political parties, motivated reasoning, and issue framing effects. *The Journal of Politics*, *72*(3), 630–645. https://doi.org/10.1017/S002238161000006X

Sniderman, P. M., Brody, R. A., & Tetlock, P. E. (1991). *Reasoning and choice: Explorations in political psychology*. Cambridge University Press.

Sternthal, B., Dholakia, R., & Leavitt, C. (1978). The persuasive effect of source credibility: Tests of cognitive response. *Journal of Consumer Research*, *4*(4), 252–260. https://doi.org/10.1086/208704

Sülflow, M., Schäfer, S., & Winter, S. (2019). Selective attention in the news feed: An eye-tracking study on the perception and selection of political news posts on Facebook. *New Media & Society*, *21*(1), 168–190. https://doi.org/10.1177/1461444818791520

Swire, B., Berinsky, A. J., Lewandowsky, S., & Ecker, U. K. H. (2017). Processing political misinformation: Comprehending the Trump phenomenon. *Royal Society Open Science*, *4*(3), 160802. https://doi.org/10.1098/rsos.160802

Taber, C. S., Cann, D., & Kucsova, S. (2009). The motivated processing of political arguments. *Political Behavior*, *31*(2), 137–155. https://doi.org/10.1007/s11109-008-9075-8

Taber, C. S., & Lodge, M. (2006). Motivated skepticism in the evaluation of political beliefs. *American Journal of Political Science*, *50*(3), 157–184. https://doi.org/10.1111/j.1540-5907.2006.00214.x

Tappin, B. M., Pennycook, G., & Rand, D. G. (2018). *Rethinking the Link Between Cognitive Sophistication and Identity-Protective Bias in Political Belief Formation*, (pp.1-60). https://doi.org/10.31234/osf.io/yuzfj

Tibshirani, J., Wager, S., & Athey, S. (2020). *The Grf algorithm*. https://grf-labs.github.io/grf/REFERENCE.html

Trilling, D., Tolochko, P., & Burscher, B. (2017). From newsworthiness to shareworthiness: How to predict news sharing based on article characteristics. *Journalism & Mass Communication Quarterly*, *94*(1), 38–60. https://doi.org/10.1177/1077699016654682

Tsfati, Y. (2010). Online news exposure and trust in the mainstream media: Exploring possible associations. *American Behavioral Scientist*, *54*(1), 22–42. https://doi.org/10.1177/0002764210376309

Tsfati, Y., & Ariely, G. (2014). Individual and contextual correlates of trust in media across 44 countries. *Communication Research*, *41*(6), 760–782. https://doi.org/10.1177/0093650213485972

Tsfati, Y., & Cappella, J. N. (2005). Why do people watch news they do not trust? The need for cognition as a moderator in the association between news media skepticism and exposure. *Media Psychology*, *7*(3), 251–271. https://doi.org/10.1207/S1532785XMEP0703_2

Tucker, J. A., Theocharis, Y., Roberts, M. E., & Barberá, P. (2017). From liberation to turmoil: Social media and democracy. *Journal of Democracy*, *28*(4), 46–59. https://doi.org/10.1353/jod.2017.0064

Valenzuela, S., Piña, M., & Ramírez, J. (2017). Behavioral effects of framing on social media users: How conflict, economic, human interest, and morality frames drive news sharing. *Journal of Communication*, *67*(5), 803–826. https://doi.org/10.1111/jcom.12325

Vallone, R. P., Ross, L., & Lepper, M. R. (1985). The hostile media phenomenon. Biased perception and perceptions of media bias in coverage of the Beirut Massacre. *Journal of Personality and Social Psychology*, *49*(3), 577–585. https://doi.org/10.1037/0022-3514.49.3.577

Van Aelst, P., Strömbäck, J., Aalberg, T., Esser, F., de Vreese, C., Matthes, J., Hopmann, D., Salgado, S., Hubé, N., Stępińska, A., Papathanassopoulos, S., Berganza, R., Legnante, G., Reinemann, C., Sheafer, T., & Stanyer, J. (2017). Political communication in a high-choice media environment: A challenge for democracy? *Annals of the International Communication Association*, *41*(1), 3–27. https://doi.org/10.1080/23808985.2017.1288551

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359* (6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Wager, S., & Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, *113*(523), 1228–1242. https://doi.org/10.1080/01621459.2017.1319839

Washburn, A. N., & Skitka, L. J. (2017). Science denial across the political divide. *Social Psychological and Personality Science, 9*(8), 972-980. https://doi.org/10.1177/1948550617731500

Zaller, J. (1992). *The nature and origins of mass opinion*. Cambridge University Press.

Zimmermann, F., & Kohring, M. (2020). Mistrust, disinforming news, and vote choice: A panel survey on the origins and consequences of believing disinformation in the 2017 German parliamentary election. *Political Communication*, *37*(2), 215–237. https://doi.org/10.1080/10584609.2019.1686095