Believing and Sharing Information by Fake Sources: An Experiment

Paul C. Bauer (MZES) & Bernhard Clemm von Hohenberg (EUI)[‡]

Published version (*Political Communication*):
http://dx.doi.org/10.1080/10584609.2020.1840462

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

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

†AFFILIATIONS: Mannheim Centre for European Social Research (MZES), University Mannheim, Mannheim, Germany, European University Institute, Florence, Italy

‡Both authors contributed equally to this work. Code and data to reproduce the study are available here: https://doi.org/10.7910/DVN/EJYWCK. 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.

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.

1 Introduction

How citizens manage to stay informed about politics—or not—has been a long-standing concern of political science (Lippmann 1922; Berelson, Lazarsfeld, and McPhee 1954; Delli Carpini and Keeter 1996; Druckman 2014). 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, Johnson, and Murphy 2012; Tsfati and Ariely 2014). At the same time, the spread of misinformation is rampant (Guess, Nyhan, and Reifler 2018; Vosoughi, Roy, and Aral 2018). How do news sources affect belief and sharing of their information in this environment?

Research on sources reaches back decades (e.g., Hovland and 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 to share news reports in an online context. Previous studies on the effect of sources have mostly compared partisan outlets or politicians (Druckman 2001; Baum and Gussin 2004; 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 (Lord, Ross, and Lepper 1979; Taber and Lodge 2006; Jerit and Barabas 2012), 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 and Mutz 2011; Iyengar and Hahn 2009; G. Pennycook and Rand 2019a). 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 and 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, political knowledge, and income. 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.

2 Theory and hypotheses

2.1 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 and Weiss 1951; Hovland, Janis, and Kelley 1953). Political science joined the debate about "source cues" some decades later (Zaller 1992; Sniderman, Brody, and Tetlock 1991; Popkin 1991; Page, Shapiro, and Dempsey 1987; Mondak 1993). 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., Pennycook and Rand 2017; Swire et al. 2017; Oeldorf-Hirsch and DeVoss 2019).

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., Sternthal, Dholakia, and Leavitt 1978; Petty, Cacioppo, and Goldman 1981; Chebat et al. 1988; Greer 2003). Another manipulation concerned the source's physical *attractiveness* or likability (e.g Chaiken 1980; Mills and Aronson 1965). *Similarity*, e.g., in social or political identity, between source and receiver was another effective treatment (e.g., Kuklinski and Hurley 1994; Baum and Groeling 2009; Kuru, Pasek, and Traugott 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

[1] There is some confusion about how the concepts of "media trust" and "source credibility" relate (Kohring and 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.

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 and Rand 2017; G. Pennycook and Rand 2019b, 2019c). 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 (G. Pennycook and Rand 2019a; Epstein, Pennycook, and Rand 2019). 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 email. Researchers are only beginning to understand what makes news "shareworthy" (Boczkowski and Mitchelstein 2012; Trilling, Tolochko, and Burscher 2017; Barberá et al. 2015). 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 and Vaccari 2019) or because of a "need for chaos" (Petersen, Osmundsen, and Arceneaux 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 and Rand 2017). 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.

2.2 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 and 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 pre-existing attitudes, values, or beliefs (Lord, Ross, and Lepper 1979; Taber and Lodge 2006; MacCoun and Paletz 2009; Kahan et al. 2017). In particular, they are more likely to accept attitudinally congruent than incongruent *factual* information, irrespective of the actual truth value (Washburn and Skitka 2017; G. Pennycook and Rand 2019b).[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 (Koehler 1993; Gentzkow and Shapiro 2006). 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).*

[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 and Green 1999; Kahan 2016b; Tappin, Pennycook, and Rand 2018).

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 judgement. 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).*

2.3 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, 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 and Walejko 2008; Loges and 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; YouGov 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, Pennycook, and Rand 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 and Metzger 2000; Kiousis 2001; Kohring and Matthes 2007; Tsfati 2010; Gallup 2019). 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 (Greer 2003; Allcott and Gentzkow 2017).

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

3 Methods: design, data, measures and models

3.1 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.

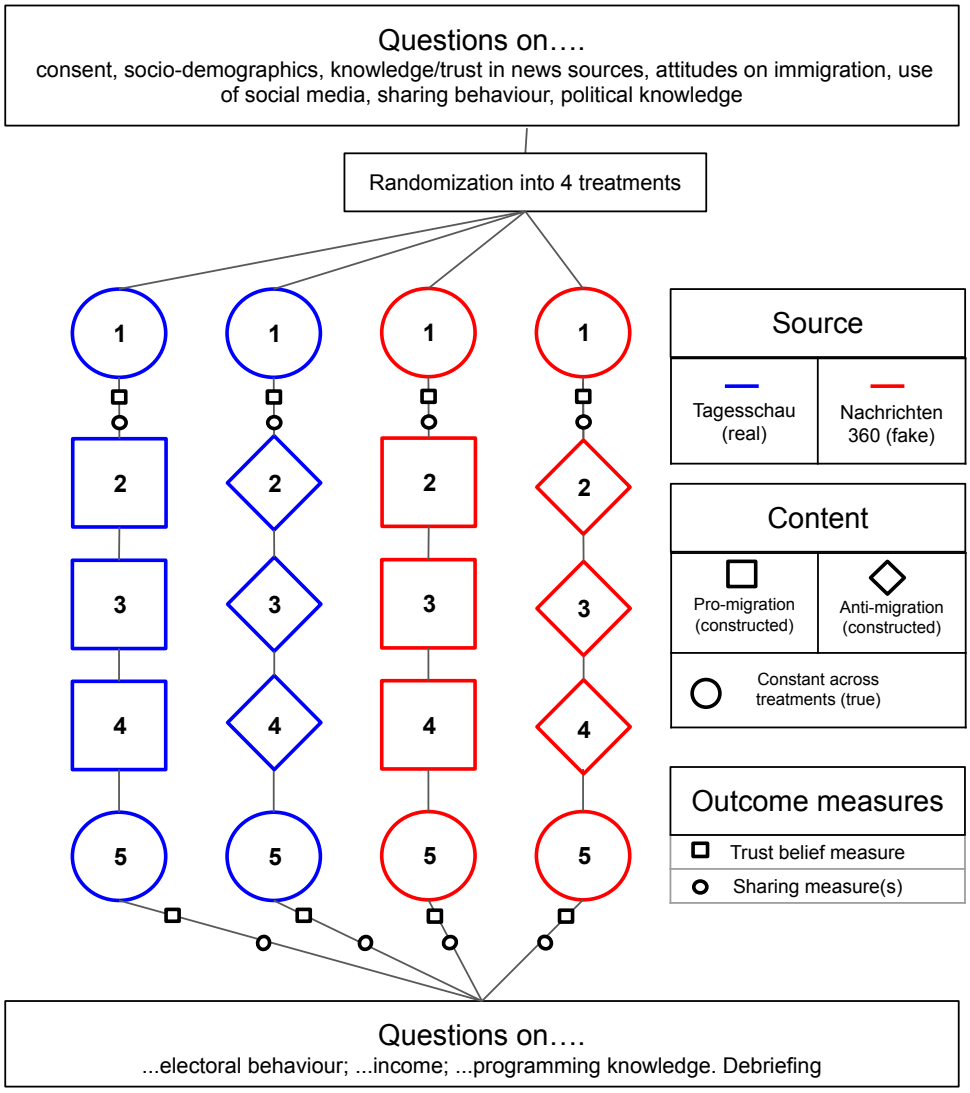
3.2 Experimental design

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 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 email, Facebook, Twitter and Whatsapp (see Figure A7) and how frequently they shared news via each platform. Sharing daily or several times a week was most common via Whatsapp (28%), followed by Twitter (19.2%), Facebook (17.9%) and email (14.9%). This question was followed by a battery measuring political knowledge and a question eliciting overconfidence in that knowledge.

Figure 1: Experimental setup



Note: 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)

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

[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

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 we measured our outcomes on: 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.

Figure 2: Example of Facebook post screenshot accompanying reports

Note: 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 percent of all crime suspects...".

We manipulated the reports along three dimensions. In the following, we refer exclusively to

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.

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 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 *Email: 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,

Pennycook, and Rand 2019). This measure of sharing intention has provided valuable insights in several recent studies (Pennycook 2020; G. Pennycook et al. 2019).

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.

3.3 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 and 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, Tibshirani, and Wager (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 by Athey, Tibshirani, and Wager (2019). See Appendix A.13.1 for details.

4 Empirical results

4.1 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, Cannon, and Rand (2018).

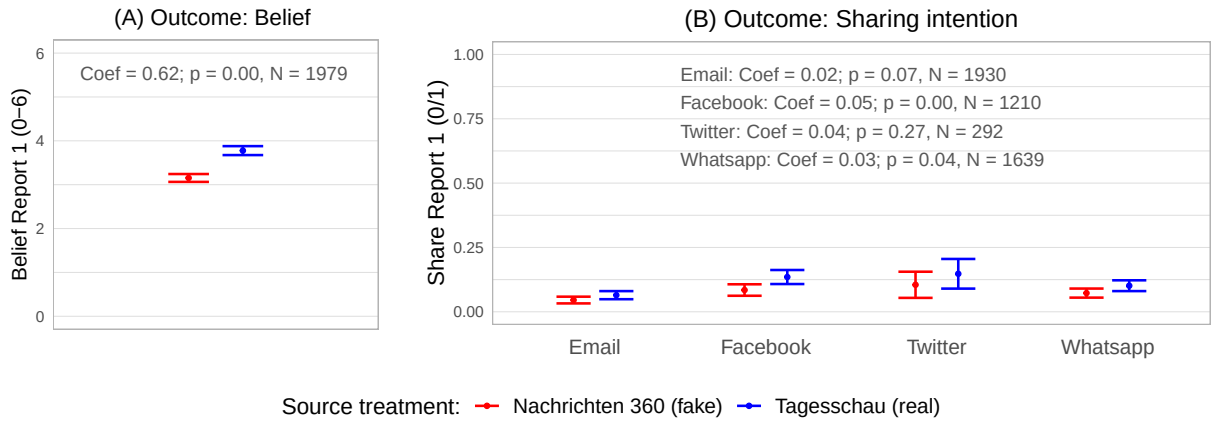
We also expected that people *would be more likely to share news reports that come from real sources than those that come from fake sources (H1b)*. Plot B in Figure 3 illustrates that intention to share is generally low. The plot also does not suggest large source effects on sharing. T-tests for email (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 correlation of 1. 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, Cannon, and Rand 2018; Pennycook 2020; G. Pennycook et al. 2019).[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, Faris, and Roberts 2018), we do not find any clear patterns in partisan sharing behavior (see Appendix A.7).

[5] Sharing intentions reported by the cited studies include both subject answering "yes" and subject answering "maybe".

Figure 3: Source effect on news belief and sharing



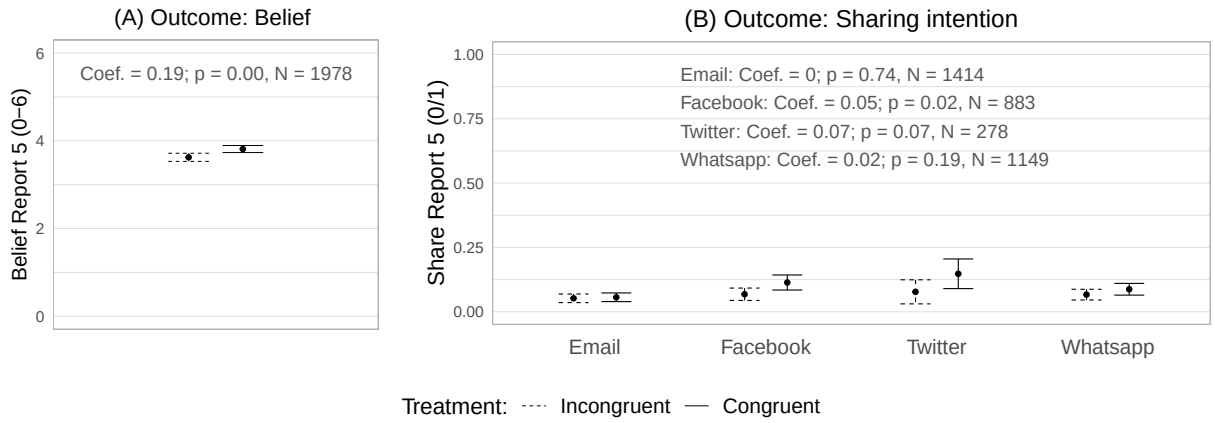
Note: Plots show point estimates and 95% confidence intervals for average belief, and sharing intentions, per source treatment group. Coefficients and p−values from t−tests are listed in the top part of the plots.

4.2 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, Pasek, and Traugott (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.

Figure 4: Congruence treatment, belief and sharing



Note: Plots show point estimates and 95% confidence intervals for average belief, and sharing intentions, per congruence treatment group. Coefficients and p−values from t−tests are listed in the top part of the plots.

There is less evidence for effects of previous congruence on later sharing (H3b). T-tests for email (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 percent 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.

4.3 Source–congruence interaction effects

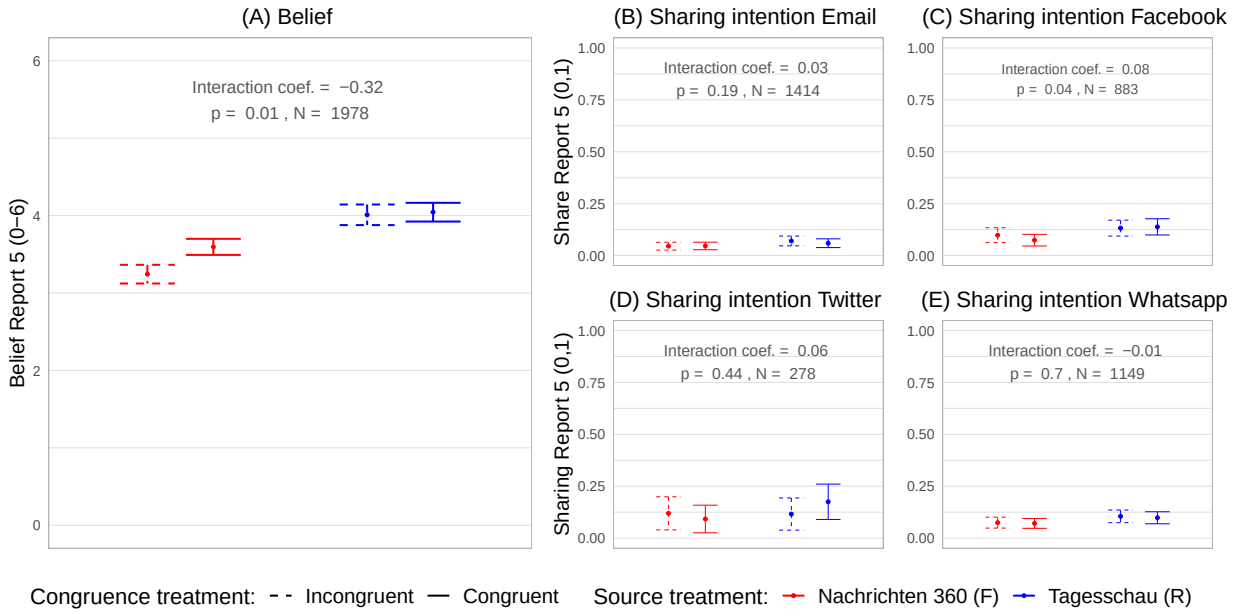
We further hypothesized that *difference in later news belief between those who saw congruent and those who saw incongruent reports would be higher for the fake source than for the real source (H3a)* and expected the same for *sharing intentions (H3b)*. Figure 5 visualizes the means in the four treatment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

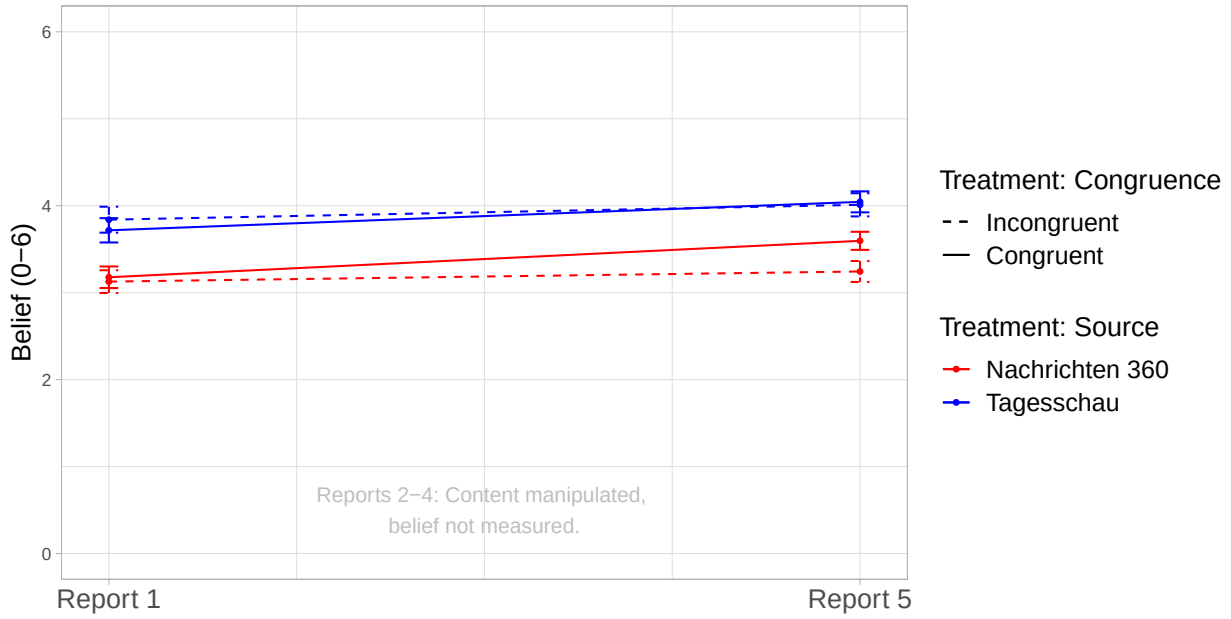
Figure 5: Interactions between source and congruence treatment



Note: Plots show point estimates and 95% confidence intervals for average belief, and sharing intentions, per source by congruence treatment group. OLS coefficients and p−values for the interaction between the two treatments listed in the top part of the plots.

[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.

Figure 6: Source credibility development



Note: Plots show point estimates and 95% confidence intervals for average belief per treatment group for both Report 1 and Report 5. Lines connect each of four treatment groups over time.

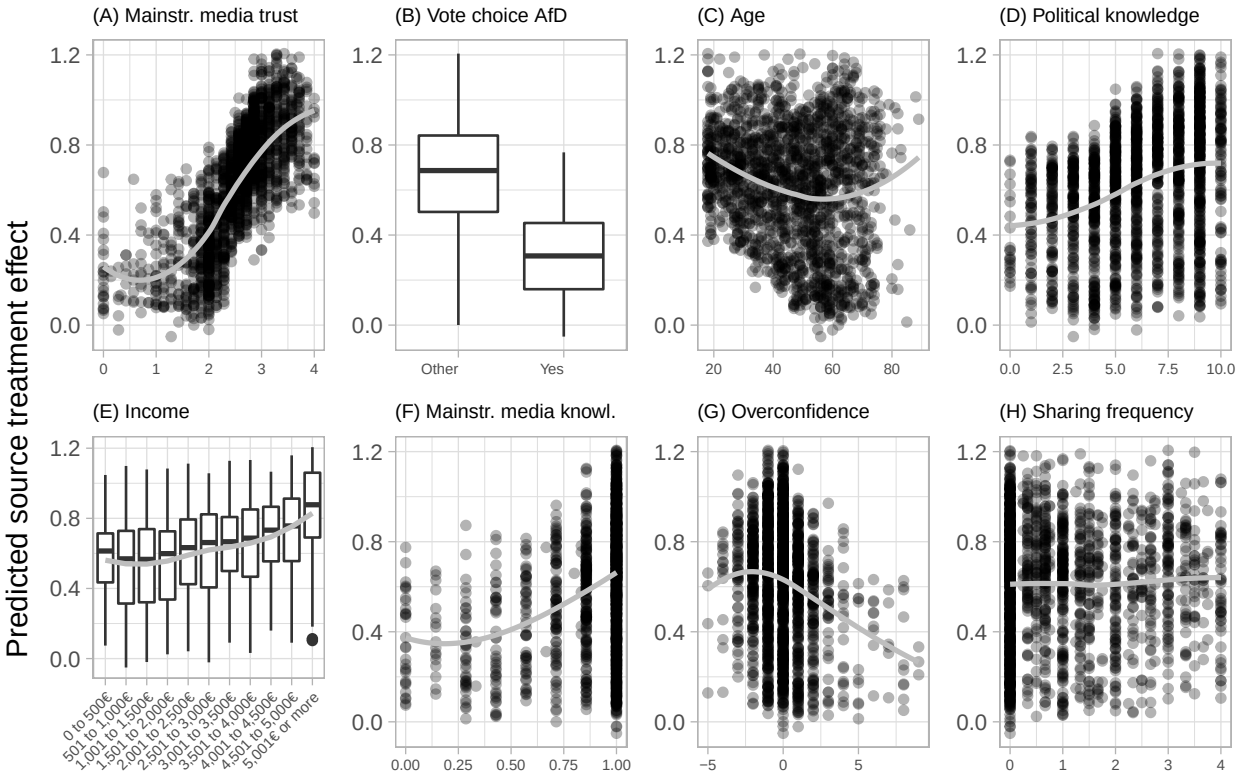
4.4 Treatment heterogeneity

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 method (cf. Wager and 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 email/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 and 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.

[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.

Figure 7: Heterogeneity of predicted source treatment effect on belief



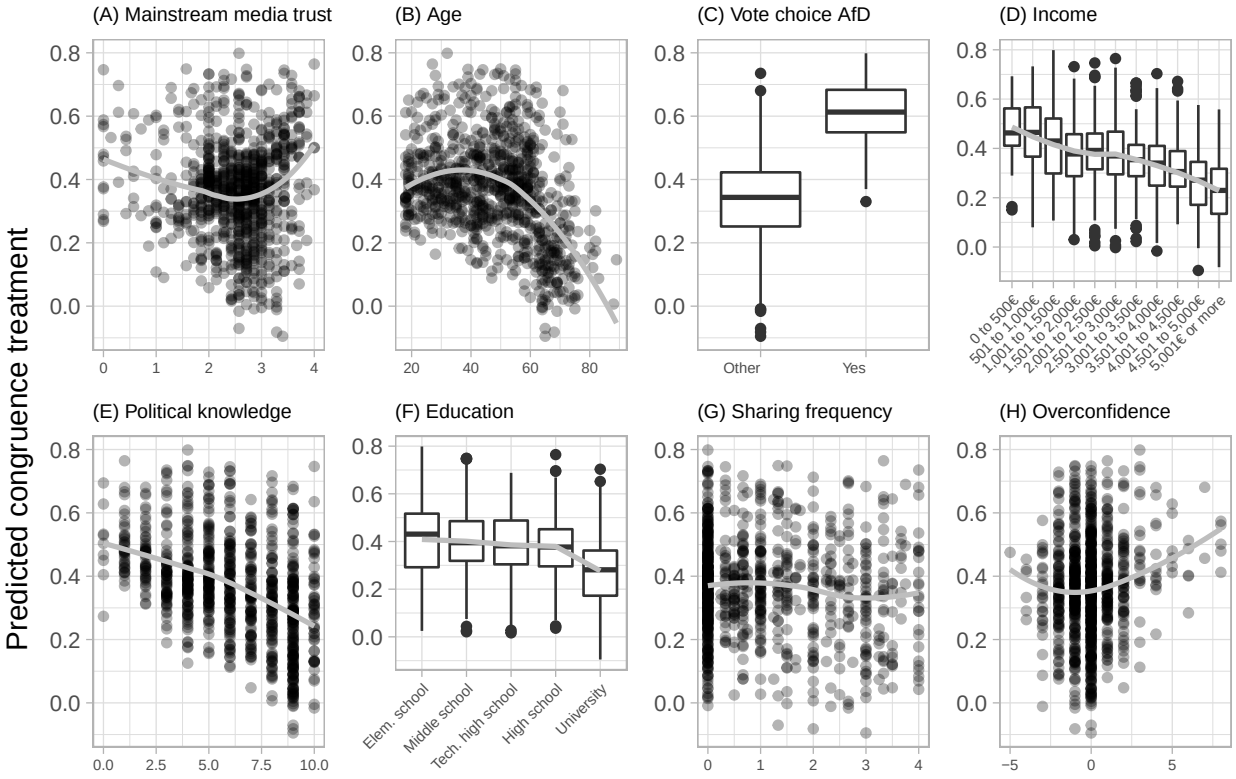
Note: Plots display variation of predicted individual treatment effects across values of continuous (scatter plots) and categorical (box plots) covariates (the 8 most important in terms of treatment heterogeneity). The gray curve is a simple smoothing curve.

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 = 0.000 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 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, Wager, and Athey 2020). It shows that only three variables are statistically significant at the 95% level: Mainstream media trust ($p = 0.01$), vote for the populist right ($p = 0.005$), and political knowledge ($p = 0.035$). Age and media knowledge show marginal significance levels (see Table A27 in the Appendix for more details).

Figure 8: Heterogeneity of predicted congruence treatment effect on belief



Note: Plots display variation of predicted individual treatment effects across values of continuous (scatter plots) and categorical (box plots) covariates (the eight most important in terms of treatment heterogeneity). The gray curve is a simple smoothing curve.

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 ($p = 0.047$).

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 predict heterogeneity. Only having voted for the populist right turns out to be statistically significant at the 95%-level ($p = 0.001$, 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.

5 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 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 and 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 (G. Pennycook et al. 2019; Bright 2016). 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., Valenzuela, Piña, and Ramírez 2017; Cappella, Kim, and Albarracín 2015). 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, Piña, and Ramírez 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, Schäfer, and Winter 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 pre-existing 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. Taber, Cann, and Kucsova 2009; Slothuus and Vreese 2010; Kahan 2016a). 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 (Tsfati and Cappella 2005; Fletcher and Park 2017) as well as with extreme attitudes (Gentzkow and Shapiro 2011; Bakshy, Messing, and Adamic 2015; Flaxman, Goel, and Rao 2016; Eady et al. 2019) 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 (A. M. Guess, Nyhan, and Reifler 2020). 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 and 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 (A. M. Guess et al. 2020). 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.

6 References

Allcott, Hunt, and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31 (2): 211–36.

American Press Institute. 2017. "'My' Media Versus 'the' Media: Trust in News Depends on Which News Media You Mean."

Arceneaux, Kevin, Martin Johnson, and Chad Murphy. 2012. "Polarized Political Communication, Oppositional Media Hostility, and Selective Exposure." *Journal of Politics* 74 (1): 174–86.

Athey, Susan, and Guido Imbens. 2016. "Recursive Partitioning for Heterogeneous Causal Effects." *PNAS* 113 (27): 7353–60.

Athey, Susan, and Guido W Imbens. 2017. "The Econometrics of Randomized Experiments." In *Handbook of Economic Field Experiments*, 1:73–140. Elsevier.

Athey, Susan, Julie Tibshirani, and Stefan Wager. 2019. "Generalized Random Forests." *Ann. Stat.* 47 (2): 1148–78.

Athey, Susan, and Stefan Wager. 2019. "Estimating Treatment Effects with Causal Forests: An Application." http://arxiv.org/abs/1902.07409.

Bakshy, Eytan, Solomon Messing, and Lada A Adamic. 2015. "Exposure to Ideologically Diverse News and Opinion on Facebook." *Science* 348 (6239): 1130–2. http://arxiv.org/abs/9809069v1.

Barberá, Pablo, Ning Wang, Richard Bonneau, John T Jost, Jonathan Nagler, Joshua A. Tucker, and Sandra González-Bailón. 2015. "The Critical Periphery in the Growth of Social Protests." *PLoS One* 10 (11): e0143611.

Baum, Matthew A., and Tim Groeling. 2009. "Partisan Cues and Public Opinion Regarding National Security and War." *Political Behavior* 31 (2): 157–86. http://arxiv.org/abs/arXiv:1011.1669v3.

Baum, Matthew A., and Phil Gussin. 2004. "In the Eye of the Beholder: How Information Shortcuts Shape Individual Perceptions of Bias in the Media."

Benkler, Yochai, Robert Faris, and Hal Roberts. 2018. *Network Propaganda: Manipulation, Disinformation, and Radicalization in American Politics*. Oxford University Press.

Berelson, Bernard R., Paul Felix Lazarsfeld, and William N. McPhee. 1954. *Voting: A Study of Opinion Formation in a Presidential Campaign.* Chicago: University of Chicago Press.

Berghel, H. 2017. "Lies, Damn Lies, and Fake News." *Computer* 50 (2): 80–85.

Boczkowski, Pablo J., and Eugenia Mitchelstein. 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.

Bright, Jonathan. 2016. "The Social News Gap: How News Reading and News Sharing Diverge." *J. Commun.* 66 (3): 343–65.

Cappella, Joseph N, Hyun Suk Kim, and Dolores Albarracín. 2015. "Selection and Transmission Processes for Information in the Emerging Media Environment: Psychological Motives and Message Characteristics." *Media Psychol.* 18 (3): 396–424.

Chadwick, Andrew, and Cristian Vaccari. 2019. "News Sharing on Uk Social Media: Misinformation, Disinformation, and Correction."

Chaiken, Shelly. 1980. "Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion." *Journal of Personality and Social Psychology* 39 (5): 752–66.

Chebat, Jean-Charles, Pierre Filiatrault, Michel Laroche, and Catherine Watson. 1988. "Compensatory Effects of Cognitive Characteristics of the Source, the Message, and the Receiver Upon Attitude Change." *Journal of Psychology;* 122 (6): 609–21.

Delli Carpini, Michael X., and Scott Keeter. 1996. *What Americans Know About Politics and Why It Matters*. New Haven: Yale University Press.

Del Vicario, Michela, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi. 2016. "The Spreading of Misinformation Online." *Proceedings of the National Academy of Sciences* 113 (3).

Ditto, Peter H., Brittany S. Liu, Cory J. Clark, Sean P. Wojcik, Eric E. Chen, Rebecca H. Grady, Jared B. Celniker, and Joanne F. Zinger. 2018. "At Least Bias Is Bipartisan: A Meta-Analytic Comparison of Partisan Bias in Liberals and Conservatives." *Perspectives on Psychological Science*. http://arxiv.org/abs//doi.org/10.1007/s10551-%20015-2769-z.For.

Druckman, James N. 2014. "Pathologies of Studying Public Opinion, Political Communication, and Democratic Responsiveness." *Political Communication* 31 (3): 467–92.

Druckman, J. N. 2001. "Using Credible Advice to Overcome Framing Effects." *Journal of Law, Economics, and Organization* 17 (1): 62–82.

Eady, Gregory, Jonathan Nagler, Andrew Guess, Jan Zilinsky, and Joshua A. Tucker. 2019. "How Many People Live in Political Bubbles on Social Media? Evidence from Linked Survey and Twitter Data." *SAGE Open* 9 (1): 2158244019832705.

Epstein, Z I V, Gordon Pennycook, and David G Rand. 2019. "Will the Crowd Game the Algorithm ? Using Layperson Judgments to Combat Misinformation on Social Media by Downranking Distrusted Sources" 1 (1): 1–15.

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

Flanagin, Andrew J., and Miriam J. Metzger. 2000. "Perceptions of Internet Information Credibility." *Journalism and Mass Communication Quarterly* 77 (3): 515–40.

Flaxman, Seth R., Sharad Goel, and Justin M. Rao. 2016. "Filter Bubbles, Echo Chambers, and Online News Consumption." *Public Opinion Quarterly* 80 (Special Issue): 298–320.

Fletcher, Richard, and Sora Park. 2017. "The Impact of Trust in the News Media on Online News Consumption and Participation." *Digital Journalism* 5 (10): 1281–99.

Fogg, B J, Jonathan Marshall, Othman Laraki, Alex Osipovich, Chris Varma, Nicholas Fang, Jyoti Paul, et al. 2001. "What Makes Web Sites Credible? A Report on a Large Quantitative Study Chi 2001." In *Proceedings of the Sigchi Conference on Human Factors in Computing Systems.*, 3:61–68. 1.

Gallup. 2019. "Gallup Poll Social Series: Governance."

Gentzkow, Matthew, and Jesse M. Shapiro. 2006. "Media Bias and Reputation." *Journal of Political Economy* 114 (2).

———. 2011. "Ideological Segregation Online and Offline." *Quarterly Journal of Economics* 126 (4): 1799–1839. http://arxiv.org/abs/arXiv:1011.1669v3.

Gerber, Alan, and Donald Green. 1999. "Misperceptions About Perceptual Bias." *Annual Review of Political Science* 2: 189–210.

Gereme, Fantahun Bogale, and William Zhu. 2019. "Early Detection of Fake News 'Before It Flies High'." In *Proceedings of the 2nd International Conference on Big Data Technologies*, 142–48. ICBDT2019. New York, NY, USA: Association for Computing Machinery.

Goldman, Seth K., and Diana C. Mutz. 2011. "The Friendly Media Phenomenon: A Cross-National Analysis of Cross-Cutting Exposure." *Political Communication* 28 (1): 42–66.

Greer, Jennifer D. 2003. "Evaluating the Credibility of Online Information: A Test of Source and Advertising Influence." *Mass Communication and Society* 6 (1): 11–28.

Guess, Andrew M, Michael Lerner, Benjamin Lyons, Jacob M Montgomery, Brendan Nyhan, Jason Reifler, and Neelanjan Sircar. 2020. "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.*, June.

Guess, Andrew M, Brendan Nyhan, and Jason Reifler. 2020. "Exposure to Untrustworthy Websites in the 2016 Us Election." *Nat Hum Behav* 4 (5): 472–80.

Guess, Andrew, Brendan Nyhan, and Jason Reifler. 2018. "Selective Exposure to Misinformation: Evidence from the." *European Research Council*.

Gunther, Albert C. 1992. "Biased Press or Biased Public? Attitudes Toward Media Coverage of Social Groups." *Public Opinion Quarterly* 56 (2): 147.

Hargittai, Eszter, and Gina Walejko. 2008. "The Participation Divide: Content Creation and Sharing in the Digital Age." *Inf. Commun. Soc.* 11 (2): 239–56.

Hovland, Carl Iver, Irving Lester Janis, and Harold H. Kelley. 1953. *Communication and Persuasion*. New Haven, CT: Yale University Press.

Hovland, Carl Iver, and Walter Weiss. 1951. "The Influence of Source Credibility on Communication Effectiveness." *The Public Opinion Quarterly* 15 (4): 635–50.

Iyengar, Shanto, and Kyu S Hahn. 2009. "Red Media, Blue Media: Evidence of Ideological Selectivity in Media Use." *Journal of Communication* 59: 19–39.

Jerit, Jennifer, and Jason Barabas. 2012. "Partisan Perceptual Bias and the Information Environment." *The Journal of Politics* 74 (3): 672–84.

Jost, John T. 2017. "Ideological Asymmetries and the Essence of Political Psychology." *Political Psychology* 38 (2): 416–17.

Kahan, Dan M. 2013. "Ideology, Motivated Reasoning, and Cognitive Reflection." *Judgment and Decision Making* 8 (4): 407–24.

———. 2016a. "The Politically Motivated Reasoning Paradigm, Part 1: What Politically Motivated Reasoning Is and How to Measure It." In *Emerging Trends in Social and Behavioral Sciences*, edited by Robert Scott and Stephen Kosslyn. New York, NY: John Wiley & Sons.

———. 2016b. "The Politically Motivated Reasoning Paradigm, Part 2: Unanswered Questions." In *Emerging Trends in the Social and Behavioral Sciences*, edited by Robert Scott and Stephen Kosslyn, 1–15. New York, NY: John Wiley & Sons, Inc.

Kahan, Dan M., Ellen Peters, Erica Cantrell Dawson, and Paul Slovic. 2017. "Motivated Numeracy and Enlightened Self-Government." *Behavioural Public Policy*, Yale law school, public law working paper no. 307, 1 (1): 54–86.

Kiousis, Spiro. 2001. "Public Trust or Mistrust? Perceptions of Media Credibility in the Information Age." *Mass Communication and Society* 4 (4): 381–403.

Knight Foundation/Gallup. 2018. "Perceived Accuracy and Bias in the Media." Washington, D.C.: Knight Foundation/Gallup.

Koehler, Jonathan J. 1993. "The Influence of Prior Beliefs on Scientific Judgements of Evidence Quality." *Organizational Behavior and Human Decision Processes* 56: 28–55.

Kohring, Matthias, and J}org Matthes. 2007. "Trust in News Media. Development and Validation of a Multidimensional Scale." *Communication Quarterly* 34 (2): 231–52.

Kuklinski, James H., and Norman L. Hurley. 1994. "On Hearing and Interpreting Political Messages : A Cautionary Tale of Citizen Cue-Taking." *The Journal of Politics* 56 (3): 729–51.

Kuru, Ozan, Josh Pasek, and Michael W. Traugott. 2017. "Motivated Reasoning in the Perceived Credibility of Public Opinion Polls." *Public Opinion Quarterly* 81 (2): 422–46.

Lippmann, Walter. 1922. *Public Opinion*. Mansfield Centre, CT: Martino Publishing.

Loges, William E, and Joo-Young Jung. 2001. "Exploring the Digital Divide: Internet Connectedness and Age." *Communic. Res.* 28 (4): 536–62.

Lord, Charles G., Lee Ross, and Mark R. Lepper. 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.

MacCoun, Rob, and Susannah Paletz. 2009. "Citizens' Perceptions of Ideological Bias in Policy Research." *Political Psychology* 30 (1): 43–65.

Metzger, Miriam J., Andrew J. Flanagin, Keren Eyal, Daisy R. Lemus, and Robert McCann. 2003. "Credibility in the 21st Century: Integrating Perspectives on Source, Message, and Media Credibility in the Contemporary Media Environment." In *Communication Yearbook 27*, edited by Pamela Kalbfleisch, 293–335. Mahwah, NJ: Lawrence Erlbaum.

Mills, Judson, and Elliot Aronson. 1965. "Opinion Change as a Function of the Communicator's Attractiveness and Desire to Influence." *Journal of Personality and Social Psychology* 1 (2): 173–77.

Mondak, Jeffery 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.

Mosleh, Mohsen, Gordon Pennycook, and David Rand. 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, Anne, and Christina L. DeVoss. 2019. "Who Posted That Story? Processing Layered Sources in Facebook News Posts." *Journalism and Mass Communication Quarterly*.

Page, Benjamin I, Robert Y Shapiro, and Glenn R Dempsey. 1987. "What Moves Public Opinion?" *The American Political Science Review* 81 (1): 23–44.

Pennycook, Bear, 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/https://doi.org/10.1287/mnsc.2019.3478.

Pennycook, Gordon, Tyrone D. Cannon, and David G. Rand. 2018. "Prior Exposure Increases Perceived Accuracy of Fake News." *Journal of Experimental Psychology: General* 147 (12): 1865–80.

Pennycook, Gordon, Ziv Epstein, Mohsen Mosleh, Antonio A. Arechar, Dean Eckles, and David G. Rand. 2019. "Understanding and Reducing the Spread of Misinformation Online."

Pennycook, Gordon, and David G Rand. 2017. "Assessing the Effect of Disputed Warnings and Source Salience on Perceptions of Fake News Accuracy," 1–51.

Pennycook, Gordon, and David G. Rand. 2019a. "Who Falls for Fake News? The Roles of Bullshit Receptivity, Overclaiming, Familiarity, and Analytic Thinking." *Journal of Personality*, 1–16.

———. 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.

Petersen, Michael Bang, Mathias Osmundsen, and Kevin Arceneaux. 2018. "A need for Chaos and the Sharing of Hostile Political Rumors in Advanced Democracies."

Petty, Richard E., John T. Cacioppo, and Rachel Goldman. 1981. "Personal Involvement as a Determinant of Argument-Based Persuasion." *Journal of Personality and Social Psychology* 41 (5): 847–55.

Popkin, Samuel L. 1991. *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns*. Chicago: University of Chicago Press.

Pornpitakpan, Chanthika. 2004. "The Persuasiveness of Source Credibility : A Critical Review of Five Decades' Evidence." *Journal of Applied Social Psychology* 34 (2): 243–81.

PricewaterhouseCoopers. 2018. "Vertrauen in Medien 2018."

Schulz, Anne, Philipp Mu, Christian Schemer, Dominique Stefanie Wirz, Martin Wettstein, and Werner Wirth. 2018. "Measuring Populist Attitudes on Three Dimensions." *International Journal of Public Opinion Research* 30 (2).

Seidman, Gwendolyn. 2013. "Self-Presentation and Belonging on Facebook: How Personality Influences Social Media Use and Motivations." *Personality and Individual Differences* 54 (3): 402–7.

Silverman, Craig. 2016. "This Analysis Shows How Viral Fake Election News Stories Outperformed Real News on Facebook."

Slothuus, Rune, and Claes H de Vreese. 2010. "Political Parties, Motivated Reasoning, and Issue Framing Effects." *J. Polit.* 72 (3): 630–45.

Sniderman, Paul M ., Richard A. Brody, and Philip E. Tetlock. 1991. *Reasoning and Choice: Explorations in Political Psychology.* New York: Cambridge University Press.

Sternthal, Brian, Ruby Dholakia, and Clark Leavitt. 1978. "The Persuasive Effect of Source Credibility: Tests of Cognitive Response." *Journal of Consumer Research* 4: 252–60.

Sülflow, Michael, Svenja Schäfer, and Stephan Winter. 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): 168–90.

Swire, Briony, Adam J. Berinsky, Stephan Lewandowsky, and Ullrich K. H. Ecker. 2017. "Processing Political Misinformation: Comprehending the Trump Phenomenon." *Royal Society Open Science* 4.

Taber, Charles S, Damon Cann, and Simona Kucsova. 2009. "The Motivated Processing of Political Arguments." *Political Behavior* 31 (2): 137–55.

Taber, Charles S., and Milton Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50 (3): 157–84.

Tappin, Ben M., Gordon Pennycook, and David G Rand. 2018. "Rethinking the Link Between Cognitive Sophistication and Identity-Protective Bias in Political Belief Formation."

Tibshirani, Julie, Stefan Wager, and Susan Athey. 2020. "The Grf Algorithm." https://grf-labs.github.io/grf/REFERENCE.html.

Trilling, Damian, Petro Tolochko, and Bj}orn Burscher. 2017. "From Newsworthiness to Shareworthiness: How to Predict News Sharing Based on Article Characteristics." *Journalism and Mass Communication Quarterly* 94 (1): 38–60.

Tsfati, Yariv. 2010. "Online News Exposure and Trust in the Mainstream Media: Exploring Possible Associations." *American Behavioral Scientist* 54 (1): 22–42.

Tsfati, Yariv, and Gal Ariely. 2014. "Individual and Contextual Correlates of Trust in Media Across 44 Countries." *Communication Research* 41 (6): 760–82.

Tsfati, Yariv, and Joseph N Cappella. 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 Psychol.* 7 (3): 251–71.

Tucker, Joshua A., Yannis Theocharis, Margaret E. Roberts, and Pablo Barberá. 2017. "From Liberation to Turmoil: Social Media and Democracy." *Journal of Democracy* 28 (4): 46–59.

Valenzuela, Sebastián, Martina Piña, and Josefina Ramírez. 2017. "Behavioral Effects of Framing on Social Media Users: How Conflict, Economic, Human Interest, and Morality Frames Drive News Sharing." *J. Commun.* 67 (5): 803–26.

Vallone, Robert P., Lee Ross, and Mark R. Lepper. 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): 577–85. http://arxiv.org/abs/arXiv:1011.1669v3.

Van Aelst, Peter, Jesper Strömbäck, Toril Aalberg, Frank Esser, Claes de Vreese, Jörg Matthes, David Hopmann, et al. 2017. "Political Communication in a High-Choice Media Environment: A Challenge for Democracy?" *Annals of the International Communication Association* 41 (1): 3–27.

Vosoughi, Soroush, Deb Roy, and Sinan Aral. 2018. "The Spread of True and False News Online." *Science* 359 (6380): 1146–51.

Wager, Stefan, and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests." *J. Am. Stat. Assoc.* 113 (523): 1228–42.

Washburn, Anthony N., and Linda J. Skitka. 2017. "Science Denial Across the Political Divide." *Social Psychological and Personality Science*.

YouGov. 2017. "Reuters 2017."

Zaller, John. 1992. *The Nature and Origins of Mass Opinion*. Cambridge: Cambridge University Press.

Zimmermann, Fabian, and Matthias Kohring. 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): 215–37.

Online Appendix

Believing and Sharing Information by Fake Sources: An Experiment

Contents

A.1 Measures: Wording and coding of survey questions 2
A.2 News reports used in the survey experiment 4
A.3 Descriptive graphs and summary statistics 10
A.4 Balance statistics and randomization 14
A.5 Representativeness: Sample and population 16
A.6 Source knowledge and trust 17
A.7 Believing and sharing by partisanship 18
A.8 Believing and sharing 20
A.9 Models 21
A.10 Robustness checks 22
A.11 Ethical considerations and open feedback 31
A.12 Some screenshots from the survey 33
A.13 Treatment heterogeneity 37
A.14 Saliency of topics 42
A.15 R session info 43
A.16 Pregistration 45
A.17 References 45

A.1 Measures: Wording and coding of survey questions

The table below outlines question wording (translated from German) and coding decisions. All coding is made transparent in the R Markdown file accompanying the study.

Table A1: Measures: Wording and coding of survey questions

Variable	Question	Choices/Coding
Age	"How old are you?"	Drop-down input
Gender	"Which gender do you have?"	"Female"/"Male"
State of residence	"In which federal state do you live?"	Choice among 16 federal states
Source knowledge	"Do you recognize the following media?"	"Yes"/"No" battery for 11 news sources in random order; average index calculated for 7 mainstream sources
Source reading/watching	"And have you, whether offline or offline, ever read or viewed news of the respective medium?"	"Yes"/"No" battery for 11 news sources in random order; Asked only news sources the respondents indicated recognizing
Source trust	"Even if you don't know all of them: Do you think you can trust the following media?"	Labelled 5-point scale: "Not at all", "rather not", "partly", "rather", "completely"; average index calculated for 7 mainstream sources
Immigration culture attitude	"Would you say that cultural life in Germany is generally undermined or enriched by immigrants? 0 means undermined, 10 means enriched."	11-point scale with labelled end points; average index calculated together with other immigration attitude items
Immigration economy attitude	"Would you say that it is generally bad or good for the German economy that immigrants come here? 0 means bad, 10 means good."	11-point scale with labelled end points
Immigration security attitude	"Would you say that Germany becomes less safe or safer with immigrants? 0 means less safe and 10 means safer."	11-point scale with labelled end point s
Immigration life attitude	"Would you say that immigrants make Germany a worse or better place to live? 0 means worse and 10 means better."	11-point scale with labelled end points
Immigration border attitude	"Would you say that Europe sufficiently protects its external borders against illegal immigration? 0 means protection is not sufficient, 10 means protection is sufficient."	11-point scale with labelled end points
Platform use	"Do you have an Email/Facebook/Twitter account? Do you have Whatsapp installed?"	Choice between Yes/Yes, but I don't use it/No
Sharing frequency	"How often do you spread news reports via Email/Facebook/Twitter/Whatsapp?"	Choice between "Less"/"Several times a year"/"Several times a month"/"Once a week"/"Daily"
Political knowledge	"Which is the party of the following politician?"	Choice between 7 parties and "Don't know" asked for 9 politicians; combined into index adding correct answers
Overconfidence	"You just answered nine knowledge questions. How many do you think you answered correctly?"	10-point scale from 0 to 9; overconfidence computed by subtracting knowledge index from this answer
Education	"What is the highest educational attainment you have achieved?"	Choice between 7 options
Turnout	"Have you voted at the 2017 federal elections?"	Choice between "Yes"/"No"/"I was not eligible to vote"/"Don't Know"
Vote/Hypothetical vote	"Which party have you voted for with your main vote ("Zweitstimme")?" If respondent had not voted or did not know: "Which party would you have voted with your main vote ("Zweitstimme")?"	Choice between six main parties/"Other party"/"Don't Know"

Variable	Question	Choices/Coding
Income	What is the complete net income of your household?	Choice between 11 income brackets, from “0 - 500 Euro” to “5001 Euro or more”

A.2 News reports used in the survey experiment

Below, we provide the texts of the five experimental reports in the original German, and a translated English version. Reports 1 and 5 were true but slightly shortened reports and not manipulated, so the only text variation is the reference to the source. Reports 2–4 were manipulated both with regard to their content and the source. Differences are indicated in square brackets, with pro-migration contents first and anti-migration contents second.

A.2.1 Report 1

A.2.1.1 Original German version

Headline: Weniger Anträge als erwartet

Teaser (in Screenshot only): Die Frage des Familiennachzugs hatte die Koalitionsverhandlungen schwer belastet. Nun zeigt sich: Nur wenige Flüchtlinge wollen ihre Angehörigen nachholen.

Text: Flüchtlinge mit eingeschränktem Schutz wollen viel weniger Familienangehörige nach Deutschland nachziehen lassen als vielfach prognostiziert. Von Anfang August bis zum 5. November 2018 sind lediglich 786 Visa für den Familiennachzug zu subsidiär Schutzberechtigten erteilt worden. Das geht aus einem Vermerk der Bundesregierung hervor, der [SOURCE] vorliegt. Von August bis Jahresende sind maximal 5000 Visa möglich.

Die Frage des Familiennachzugs zu den etwa 300.000 Flüchtlingen, die mit sogenanntem subsidiären Schutzstatus in Deutschland leben, hatte zu Jahresbeginn die Koalitionsverhandlungen zwischen Union und SPD schwer belastet. Politiker hatten gewarnt, dass bis zu 300.000 Angehörige nach Deutschland kommen wollten, sollte der damals geltende Stopp des Familiennachzugs auslaufen.

Die Koalition einigte sich auf den Kompromiss, monatlich bis zu 1000 Visa für enge Angehörige - also Ehepartner, minderjährige Kinder oder Eltern hier lebender Minderjähriger - zu vergeben. Doch diese Zahl wird auch nach Einschätzung des Innenministeriums kaum erreicht werden, wie [SOURCE] aus Ministeriumskreisen erfahren hat.

A.2.1.2 Translated English version

Headline: Fewer applications than expected

Teaser (in screenshot only): The issue of family reunification had put a heavy strain on the coalition negotiations. Now it is clear that only a few refugees want to bring their relatives to join them.

Text: Refugees with limited protection want much fewer family members to follow them to Germany than many had predicted. Between the beginning of August and November 5, 2018, only 786 visas have been issued for family reunification with persons granted subsidiary protection. This is the result of a statement issued by the federal government, which has been published in [SOURCE]. From August to the end of the year, a maximum of 5,000 visas are possible.

At the beginning of the year, the issue of family reunification with the approximately 300,000 refugees who are living in Germany with so-called subsidiary protection status had put a heavy strain on the coalition negotiations between the CDU/CSU and the SPD. Politicians had warned that up to 300,000 relatives would want to come to Germany if the stop on family reunification, which was in effect at the time, expired.

The coalition agreed on a compromise to issue up to 1000 visas per month for close relatives - i.e. spouses, under-age children or parents of minors living here. But even according to estimates by the interior ministry this number will hardly be reached, as [SOURCE] has learned from ministry circles.

A.2.2 Report 2

A.2.2.1 Original German version *Headline:* Zuwanderer unter Tatverdächtigen [nicht überrepräsentiert / überrepräsentiert]

Teaser (in Screenshot only): Die Zahlen der neuesten Polizeistatistik zeigen: Zuwanderer sind in Deutschland nicht [überdurchschnittlich / überdurchschnittlich] oft tatverdächtig.

Text: Im vergangenen Jahr waren [1,9 / 15,1] Prozent aller Straftatverdächtigen in Deutschland Zuwanderinnen und Zuwanderer. Insgesamt wurden 1,97 Millionen Verdächtige einer Straftat registriert, davon waren [37.233 / 297.273] Zuwanderer. Das geht aus der Polizeilichen Kriminalstatistik (PKS) für 2017 hervor, die [SOURCE] vorab analysiert hat.

Als Zuwanderer erfasst die PKS 2017 Asylbewerber, Geduldete, "unerlaubt Aufhältige", subsidiär Geschützte sowie die wenigen Kontingentflüchtlinge, die über internationale Hilfsprogramme in Deutschland Aufnahme fanden. Der so definierte Personenkreis hat Ende 2017 etwa zwei Prozent der Bevölkerung ausgemacht. Es gab vergangenes Jahr also [etwa so viele / mehr] Verdachtsfälle gegen Zuwanderer, [wie / als] es ihrem Anteil an der Gesamtbevölkerung entspricht. Bei den Berechnungen wurden jeweils ausländerrechtliche Straftaten wie etwa ein illegaler Aufenthalt ausgeklammert, da diese nicht von Deutschen begangen werden können.

Die Informationen bestätigen einen Trend, über den [SOURCE] bereits im Vorjahr berichtet hatte. Demnach sind Zuwanderer [nicht überdurchschnittlich / überdurchschnittlich] oft an der gesamten registrierten Kriminalität beteiligt.

A.2.2.2 Translated English version *Headline:* Immigrant [not overrepresented / overrepresented] among criminal suspects

Teaser (in screenshot only): The figures from the latest police statistics show: Immigrants [are / are not] suspected of a crime in Germany above average.

Text: Last year, [1.9 / 15.1] percent of all suspected criminals in Germany were immigrants. A total of 1.97 million suspects of a crime were registered, of which [37,233 / 297,273] were immigrants. This is the result of the police crime statistics (PKS) for 2017, which [SOURCE] has analyzed in advance.

The 2017 statistics include asylum seekers, tolerated persons, "unauthorised residents", people granted subsidiary protection and the few contingent refugees who have been accepted in Germany through international aid programmes. At the end of 2017, the group of people defined in this way made up about two percent of the population. So last year there were [about as many / more] suspected cases against immigrants [as / than] corresponds to their share of the total population. In the calculations, criminal offences under aliens law, such as illegal residence, were excluded, as these cannot be committed by Germans.

The information confirms a trend that [SOURCE] had already reported in the previous year. Accordingly, immigrants [are / are not] involved in the total amount of registered crime above average.

A.2.3 Report 3

A.2.3.1 Original German version *Headline:* Mehrheit von Flüchtlingen schafft den [Deutschtest / Deutschtest nicht]

Teaser (in Screenshot only): Seit zwölf Jahren gibt es Integrationskurse für Flüchtlinge – [mit / ohne] Erfolg, wie die Abschlusszahlen des letzten Jahres zeigen.

Text: Seit mehr als zwölf Jahren lernen Zuwanderer Deutsch in den Integrationskursen des Bundesamts für Migration und Flüchtlinge (BAMF). Nach [anfänglichen Schwierigkeiten sind die Kurse inzwischen ein Erfolg / anfänglichem Erfolg stecken die Kurse inzwischen in Schwierigkeiten]: Eine Auswertung der Statistiken durch [SOURCE] zeigt, dass die Abschlussquoten deutlich [gestiegen / gesunken] sind.

So nahmen 2017 genau 376.468 Menschen erstmals an einem Integrationskurs teil. In diesem Zeitraum haben [337.504 / 111.623] Teilnehmer den Integrationskurs erfolgreich absolviert, also circa [90 / 30] Prozent. Die Erfolgsquote ist damit wesentlich [höher / geringer] als noch vor wenigen Jahren: Für den Zeitraum von 2005 bis 2015 ergibt die Analyse durch [SOURCE] einen Wert von knapp 1,1 Millionen Absolventen bei über 1,8 Millionen Teilnehmern, also einer Abschlussquote von etwa 60 Prozent.

Der [Erfolg / Misserfolg] spiegelt sich auch in den Prüfungsergebnissen: Von denjenigen, die 2017 am Sprachtest zum Kursende teilnahmen, schaffte eine [Mehrheit von 71,2% / nur eine Minderheit von 31,2%] das Sprachniveau B1, dem Maßstab für ausreichende deutsche Sprachkenntnisse laut Aufenthaltsgesetz.

A.2.3.2 Translated English version *Headline*: The majority of refugees [do not pass / pass] the German test

Teaser (in screenshot only): Integration courses for refugees have been running for twelve years - [with / without] success, as the final figures of last year show.

Text: For more than twelve years, immigrants have been learning German in the integration courses of the Federal Office for Migration and Refugees (BAMF). After [initial difficulties,

the courses are now a success / initial success the courses are now in difficulties]: An evaluation of the statistics by [SOURCE] shows that the completion rates have [risen / fallen] substantially.

In 2017 exactly 376,468 people took part in an integration course for the first time. During this period [337,504 / 111,623] participants successfully completed the integration course, i.e. about [90 / 30] per cent. The success rate is thus considerably [higher / lower] than just a few years ago: For the period from 2005 to 2015, the analysis by [SOURCE] shows a figure of just under 1.1 million graduates with over 1.8 million participants, i.e. a completion rate of around 60 percent.

The [success / failure] is also reflected in the examination results: Of those who took the language test at the end of the course in 2017, a [majority of 71.2% / only a minority of 31.2%] achieved language level B1, the benchmark for sufficient German language skills according to the Residence Act.

A.2.4 Report 4

A.2.4.1 Original German version *Headline:* Es gibt [keine / eine] Sogwirkung der Seenotrettung

Teaser (in Screenshot only): Die Flüchtlingszahlen im Mittelmeer der vergangenen Jahre lassen den Schluss zu, dass die private Seenotrettung [keine / eine] Sogwirkung entfaltet.

Text: Die private Seenotrettung von Migranten zwischen Italien und Nordafrika wurde im Sommer 2018 durch die italienische Regierung so gut wie gestoppt – doch die Diskussion über die Einsätze der Retter tobt weiter. Kritiker der NGOs glauben, dass erst die Anwesenheit der Schiffe Migranten dazu bewegt, den lebensgefährlichen Trip zu wagen. Die Helfer selbst und ihre Unterstützer wollen von einer Sogwirkung ihres Handelns nichts wissen. Die Redaktion der [SOURCE] hat die Zahlen des UN-Flüchtlingshilfswerks zu den Ankünften nach Italien analysiert.

Insgesamt [entkräften / bestärken] die Daten den Verdacht, dass die Rettungsboote die Migration angekurbelt haben. Ende 2013 startete mit Mare Nostrum die erste große staatliche Rettungsmission von italienischer Seite. Im ersten Jahr der Mission [sank / stieg] die Anzahl der ankommenden Menschen gegenüber dem Vorjahr auf [85.050 / 170.000]. Die Italiener stoppten Ende 2014 das Programm. Im Jahr 2015 wurde Mare Nostrum durch die militärische EU-Mission Triton ersetzt, die mit deutlich weniger Schiffen ausgestattet war. [Trotzdem / In Folge dessen] kamen mit [153.842 / 76.921] wieder [mehr / weniger] Bootsflüchtlinge nach Italien.

Schließlich fanden sich 2016 immer mehr private Retter im Mittelmeer ein. Zur Hochzeit ihrer Aktivität [verringerte / vergrößerte] sich die Anzahl der Ankünfte, auf [59.685 / 119.369] im Jahr 2017. [Obwohl / Seit] die italienische Regierung vermehrt gegen die NGOs vorgeht, gab es 2018 wieder etwas [mehr / weniger] Ankünfte. Die von [SOURCE] analysierten Daten legen also [nicht nahe / nahe], dass die Anwesenheit der Retter eine Sogwirkung entfaltet.

A.2.4.2 Translated English version

Headline: There [is no / is a] pull effect of sea rescue

Teaser (in screenshot only): The number of refugees in the Mediterranean in recent years leads to the conclusion that private sea rescue is [not having / having a] pull effect.

Text: The private sea rescue of migrants between Italy and North Africa was almost stopped by the Italian government in the summer of 2018 - but the discussion about the rescue operations continues to rage. Critics of the NGOs believe that the presence of the ships alone will motivate migrants to make the life-threatening trip. The rescuers themselves and their supporters deny any pull effect of their actions. Reporters of [SOURCE] have analyzed the figures of the UN refugee relief organisation on the arrivals in Italy.

Overall, the data [refute / confirm] the suspicion that the lifeboats have stimulated migration. At the end of 2013 Mare Nostrum was the first major government rescue mission from Italy. In the first year of the mission the number of arrivals [sank / rose] to [85,050 / 170,000] compared to the previous year. The Italians stopped the programme at the end of 2014. In 2015 Mare Nostrum was replaced by the EU military mission Triton, which was equipped with significantly fewer ships. [Nevertheless / as a consequence] [more / less] boat refugees came to Italy again with [153.842 / 76.921].

Finally, more and more private rescuers arrived in the Mediterranean in 2016. At the peak of their activity, the number of arrivals [decreased / increased] to [59,685 / 119,369] in 2017. [Although / since] the Italian government is increasingly taking action against the NGOs, there were again [some more / fewer] arrivals in 2018. The data analyzed by [SOURCE] thus [does not suggest / suggest] that the presence of the rescuers has a pull effect.

A.2.5 Report 5

A.2.5.1 Original German version

Headline: Weniger Flüchtlinge nutzen Ausreiseförderung

Teaser (in Screenshot only): Im vergangenen Jahr haben weniger Flüchtlinge das Förderprogramm der Bundesregierung zur freiwilligen Rückkehr angenommen.

Text: Die Zahl der Flüchtlinge, die über ein Förderprogramm der Bundesregierung freiwillig in ihre Heimat zurückkehren, ist in den vergangenen drei Jahren deutlich gesunken. Von Januar bis Ende November 2018 haben 15.089 Menschen das Angebot in Anspruch genommen, wie eine Auswertung der Statistiken durch [SOURCE] ergab.

Für das gesamte Jahr 2017 ergibt die Auswertung eine Zahl von 29.522 freiwilligen Rückkehrern, 2016 waren es noch 54.006. “Die - im Vergleich relativ hohe - Zahl der bewilligten freiwilligen Ausreisen des Jahres 2016 ist im Zusammenhang mit den in diesem Zeitraum historisch hohen Zugangszahlen von in Deutschland schutzsuchenden Menschen zu sehen”, erklärte eine Sprecherin des BAMF gegenüber [SOURCE].

Die Zahlen der Behörde beziehen sich ausschließlich auf die von Bund und Ländern angebotenen Programme für Rückkehrer. Dabei werden die Reisekosten von Rückkehrwilligen

übernommen und je nach Fall auch eine weitere Reisebeihilfe und ein Startgeld für das neue Leben im Heimatland.

A.2.5.2 Translated English version

Headline: Fewer refugees use return funds

Teaser (in screenshot only): In the past year, fewer refugees have accepted the federal government's support program for voluntary return.

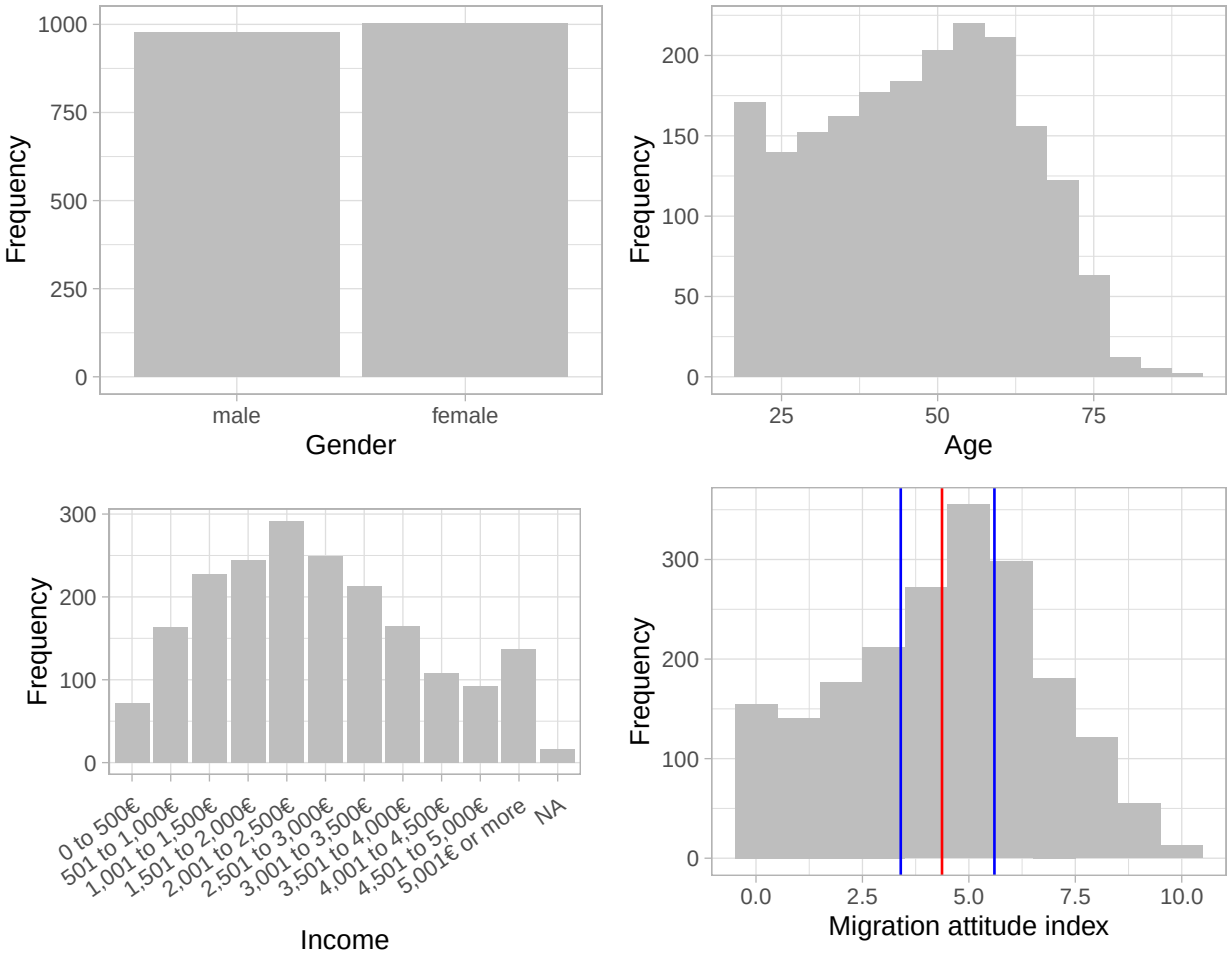
Text: The number of refugees who voluntarily return to their home country through a federal government support program has decreased significantly in the past three years. From January to the end of November 2018, 15,089 people have taken advantage of the offer, according to an evaluation of the statistics by [SOURCE].

For the entire year 2017, the evaluation reveals a number of 29,522 voluntary returnees, compared to 54,006 in 2016. "The - relatively high - number of voluntary departures granted in 2016 must be seen in connection with the historically high number of people seeking protection in Germany during this period," a spokeswoman of the BAMF told [SOURCE].

The authority's figures refer exclusively to the programmes for returnees offered by the Federal Government and the Länder. In these programmes, the travel costs of returnees are paid for and, depending on the case, also a further travel allowance and an entry fee for the new life in the home country.

A.3 Descriptive graphs and summary statistics

Figure A1: Distributions: gender, age, income and migration attitude index



Note: Plot 4 displays the distribution of the individual average across the scales measuring attitudes towards immigration. Based on these attitudes, we constructed two treatment variables in combination with the content treatment (pro− and anti−migration).

Table A2, A3, and A4 display summary statistics for all variables used in the analysis (and some more).

Table A2: Summary statistics: Numeric variables

Statistic	N	Mean	St. Dev.	Min	Max
Know Tagesschau	1,974	0.95	0.21	0.00	1.00
Know Heute	1,973	0.94	0.24	0.00	1.00
Know Sz	1,966	0.84	0.37	0.00	1.00
Know Faz	1,970	0.83	0.37	0.00	1.00
Know Focus	1,970	0.89	0.31	0.00	1.00
Know Bild	1,967	0.91	0.28	0.00	1.00
Know Nachrichten360	1,932	0.05	0.21	0.00	1.00
Know Berliner	1,941	0.21	0.41	0.00	1.00
Know Spiegel	1,972	0.91	0.28	0.00	1.00
Know Rtdeutsch	1,935	0.22	0.41	0.00	1.00
Know Newsblitz	1,934	0.04	0.20	0.00	1.00
Read Tagesschau	1,877	0.91	0.29	0.00	1.00
Read Heute	1,853	0.87	0.34	0.00	1.00
Read Sz	1,640	0.70	0.46	0.00	1.00
Read Faz	1,636	0.65	0.48	0.00	1.00
Read Focus	1,745	0.79	0.40	0.00	1.00
Read Bild	1,778	0.79	0.41	0.00	1.00
Read Nachrichten360	90	0.49	0.50	0.00	1.00
Read Berliner	406	0.41	0.49	0.00	1.00
Read Spiegel	1,798	0.84	0.36	0.00	1.00
Read Rtdeutsch	422	0.66	0.48	0.00	1.00
Read Newsblitz	79	0.41	0.49	0.00	1.00
Trust Tagesschau	1,974	2.90	1.02	0.00	4.00
Trust Heute	1,975	2.84	1.01	0.00	4.00
Trust Sz	1,971	2.66	0.93	0.00	4.00
Trust Faz	1,972	2.64	0.95	0.00	4.00
Trust Focus	1,968	2.48	0.94	0.00	4.00
Trust Bild	1,975	1.34	1.02	0.00	4.00
Trust Nachrichten360	1,967	1.69	0.86	0.00	4.00
Trust Berliner	1,968	2.06	0.87	0.00	4.00
Trust Spiegel	1,972	2.52	0.96	0.00	4.00
Trust Rtdeutsch	1,971	1.67	0.93	0.00	4.00
Trust Newsblitz	1,969	1.55	0.85	0.00	4.00
Trust Tagesschau Known	1,875	2.94	1.00	0.00	4.00
Trust Heute Known	1,852	2.88	0.98	0.00	4.00
Trust Sz Known	1,647	2.78	0.88	0.00	4.00
Trust Faz Known	1,638	2.75	0.89	0.00	4.00
Trust Focus Known	1,748	2.53	0.91	0.00	4.00
Trust Bild Known	1,787	1.39	1.02	0.00	4.00
Trust Nachrichten360 Known	93	2.41	0.84	0.00	4.00
Trust Berliner Known	407	2.34	0.86	0.00	4.00
Trust Spiegel Known	1,795	2.58	0.94	0.00	4.00
Trust Rtdeutsch Known	424	1.89	1.13	0.00	4.00
Trust Newsblitz Known	80	2.09	0.90	0.00	4.00

Table A3: Summary statistics: Numeric variables

Statistic	N	Mean	St. Dev.	Min	Max
Age	1,980	46.53	16.07	18	89
Sex	1,980	0.49	0.50	0	1
Treatment Source	1,980	0.49	0.50	0	1
Treatment Channel	1,980	0.50	0.50	0	1
Treatment Content	1,980	0.49	0.50	0	1
Treatment Congruence50	1,980	0.52	0.50	0	1
Treatment Congruence 30	1,353	0.52	0.50	0.00	1.00
Belief Report 1	1,979	3.46	1.57	0.00	6.00
Belief Report 5	1,978	3.72	1.39	0.00	6.00
Share Report 1 Email	1,930	0.05	0.23	0.00	1.00
Share Report 1 Fb	1,210	0.11	0.31	0.00	1.00
Share Report 1 Twitter	292	0.13	0.33	0.00	1.00
Share Report 1 Whatsapp	1,639	0.09	0.28	0.00	1.00
Share Report 2 Email	1,931	0.06	0.24	0.00	1.00
Share Report 2 Fb	1,210	0.11	0.31	0.00	1.00
Share Report 2 Twitter	292	0.14	0.35	0.00	1.00
Share Report 2 Whatsapp	1,641	0.11	0.31	0.00	1.00
Share Report 5 Email	1,414	0.05	0.23	0.00	1.00
Share Report 5 Fb	883	0.09	0.29	0.00	1.00
Share Report 5 Twitter	278	0.12	0.32	0.00	1.00
Share Report 5 Whatsapp	1,149	0.08	0.27	0.00	1.00
Immigrant Culture	1,980	5.01	2.90	0	10
Immigrant Economy	1,980	5.48	2.67	0	10
Immigrant Security	1,980	3.37	2.25	0	10
Immigrant Life	1,980	4.40	2.63	0	10
Immigrant Border	1,980	3.59	2.94	0	10
Sharing Email	1,930	0.79	1.29	0.00	4.00
Sharing Fb	1,208	1.05	1.32	0.00	4.00
Sharing Twitter	292	1.10	1.34	0.00	4.00
Sharing Whatsapp	1,641	1.40	1.55	0.00	4.00

Table A4: Summary statistics: Categorical variables

Variable	Distribution
Education (German school types)	
None	3 (0)
Finished Grundschule	15 (1)
Finished Volksschule or Hauptschule	257 (13)
Mittlere Reife, Realschulabschluss, Fachoberschulreife	673 (34)
Fachhochschulreife	160 (8)
Abitur	384 (19)
Fachhochschule or Hochschule	488 (25)
Income	
0 to 500€	72 (4)
501 to 1,000€	164 (8)
1,001 to 1,500€	228 (12)
1,501 to 2,000€	244 (12)
2,001 to 2,500€	292 (15)
2,501 to 3,000€	249 (13)
3,001 to 3,500€	213 (11)
3,501 to 4,000€	165 (8)
4,001 to 4,500€	108 (5)
4,501 to 5,000€	92 (5)
5,001€ or more	137 (7)
Missings	16/1,980 (1)
Use Email	
No	28 (1)
Yes, but I dont use it	18 (1)
Yes	1,931 (98)
Missings	3/1,980 (0)
Use Facebook	
No	600 (30)
Yes, but I dont use it	166 (8)
Yes	1,210 (61)
Missings	4/1,980 (0)
Use Twitter	
No	1,520 (77)
Yes, but I dont use it	161 (8)
Yes	293 (15)
Missings	6/1,980 (0)
Use Whatsapp	
No	312 (16)
Yes, but I dont use it	24 (1)
Yes	1,641 (83)
Missings	3/1,980 (0)

A.4 Balance statistics and randomization

Table A5 and A6 provide balance statistics for content and source treatment. Table A7 provides statistics across the 2*2 treatment groups. Significance tests of group differences are done with t-tests/anova for continuous variables and chi-squared tests for categorical outcomes. None of the differences is statistically significant at a 95% level.

Table A5: Balance statistics for source treatment

	Source treatment		
Variable	Nachrichten 360	Tagesschau	Diff. p-value
Age (mean)	47.15	45.89	0.08
Migration attitude (mean)	4.29	4.45	0.15
Sex (proportion)	0.51	0.47	0.06
Vote CDU/CSU (proportion)	0.20	0.21	0.98

Table A6: Balance statistics for content treatment

	Content treatment		
Variable	Pro-migration	Anti-migration	Diff. p-value
Age (mean)	46.80	46.25	0.45
Migration attitude (mean)	4.30	4.44	0.17
Sex (proportion)	0.49	0.50	0.76
Vote CDU/CSU (proportion)	0.19	0.22	0.77

Table A7: Balance statistics for source*content treatment

Variable	Source*Content treatment				Diff. p-value
	Pro-migration/Fake	Pro-migration/Real	Anti-migration/Fake	Anti-migration/Real	
Age (mean)	47.90	45.74	46.42	46.06	0.05
Migration attitude (mean)	4.25	4.34	4.34	4.57	0.05
Sex (proportion)	0.50	0.48	0.53	0.46	0.11
Vote CDU/CSU (proportion)	0.20	0.19	0.21	0.23	0.89

A.5 Representativeness: Sample and population

Our sample is a non-probability quota sample. Table A8, A9 and A10 display both population and sample statistics for the quota variables.

Table A8: Gender distribution (population and sample)

Sex	Population: absolute	Population: percent	Data: absolute	Data: percent
male	40843565	49.33	975	49.39
female	41948786	50.67	999	50.61

Table A9: Residence distribution (population and sample)

State	Population: absolute	Population: percent	Data: percent	Data: absolute
Baden-Wuerttemberg	11023425	13.31	13.32	263
Bayern	12997204	15.70	15.65	309
Berlin	3613495	4.36	4.36	86
Brandenburg	2504040	3.02	3.04	60
Bremen	681032	0.82	0.81	16
Hamburg	1830584	2.21	2.23	44
Hessen	6243262	7.54	7.55	149
Mecklenburg-Vorpommern	1611119	1.95	1.98	39
Niedersachsen	7962775	9.62	9.57	189
Nordrhein-Westfalen	17912134	21.64	21.58	426
Rheinland-Pfalz	4073679	4.92	4.91	97
Saarland	994187	1.20	1.22	24
Sachsen	4081308	4.93	4.96	98
Sachsen-Anhalt	2223081	2.69	2.68	53
Schleswig-Holstein	2889821	3.49	3.50	69
Thueringen	2151205	2.60	2.63	52

Table A10: Age distribution (population and sample)

Age	Population: absolute	Population: percent	Data: percent	Data: absolute
18-30ys	12709507	18.35	20.87	412
31-43ys	13190233	19.05	21.68	428
44-56ys	16559912	23.91	27.25	538
57-69ys	13796536	19.92	22.59	446
70+	12998017	18.77	7.60	150

A.6 Source knowledge and trust

In the first part of our study, we exposed subjects to a list of nine real or made-up news sources and asked whether they knew, had read/watched and trusted each source. Below, we report source-level averages of these outcomes. Figure A2 shows that proportion of subjects affirming they recognize a source and say that they have read or watched it. It can be seen that almost everyone knows our real source (*Tagesschau*), and only few imagine knowing the source we made up (*Nachrichten 360)*.

Figure A2: Average knowledge of news sources

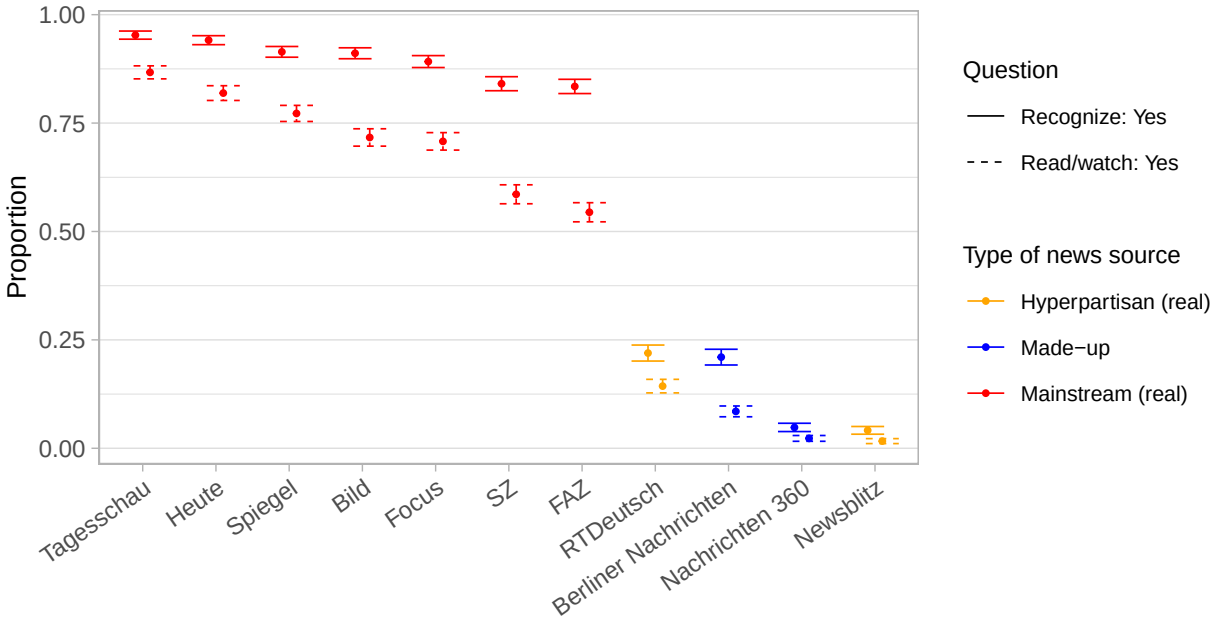
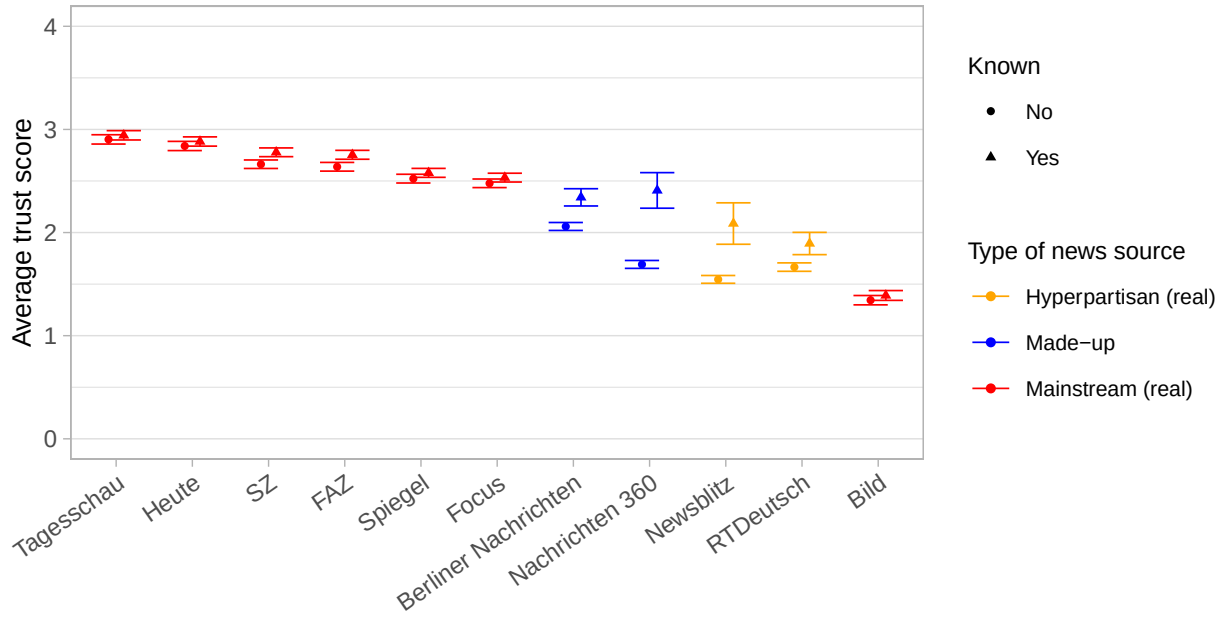


Figure A3 shows averages of subjects responses to the question: "Even if you don't know all of them: Do you think the following media can be trusted?" (Not at all, rather not, partly, rather, completely, cf. G. Pennycook and Rand 2019a), which we asked irrespective of whether people knew the source. Averages are reported separately for those who said they knew the respective source and those who said they did not.

The two Figures taken together allow some interesting interpretation of what is going on behind the source treatment manipulation: The familiarity gap between the two experimental sources is almost maximal (see Figure A2). The gap in credibility/trustworthiness perceptions is much smaller, as Figure A3 shows. This suggests that there is something else about the fake source (*Nachrichten 360*) that enhances its credibility. The fact that there is a substantive difference between the two made-up sources suggests that the name might make a difference. But (imagined) familiarity does play a role, since people who think they know the source also find it more credible than those who do not.

Figure A3: Average trust of news sources



A.7 Believing and sharing by partisanship

Figure A4 and A5 plot belief and sharing intentions against vote intention to explore whether there might be any patterns across partisans that might qualify our results.

Figure A4: Belief by vote intention

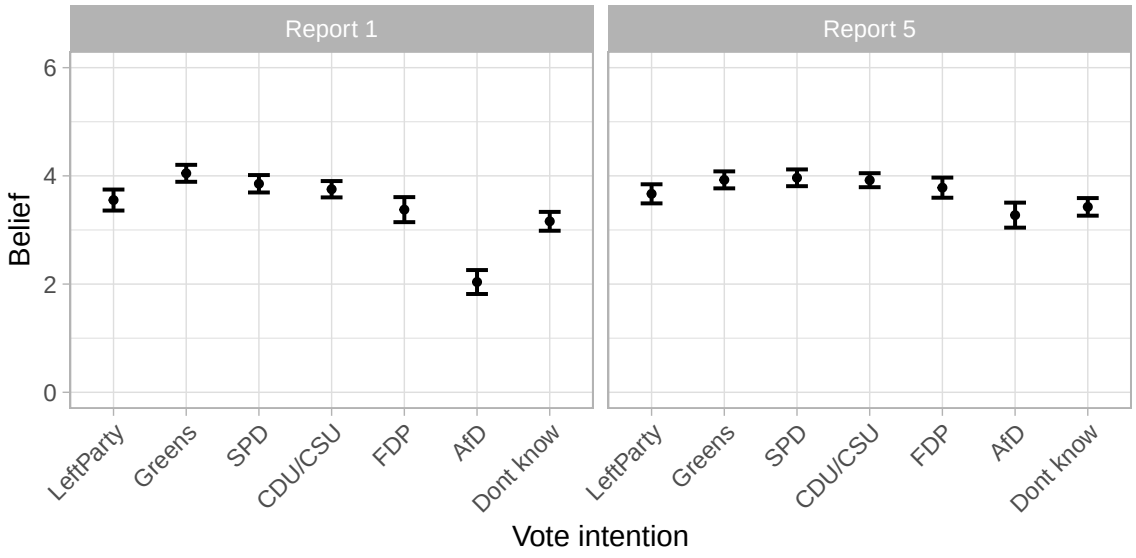
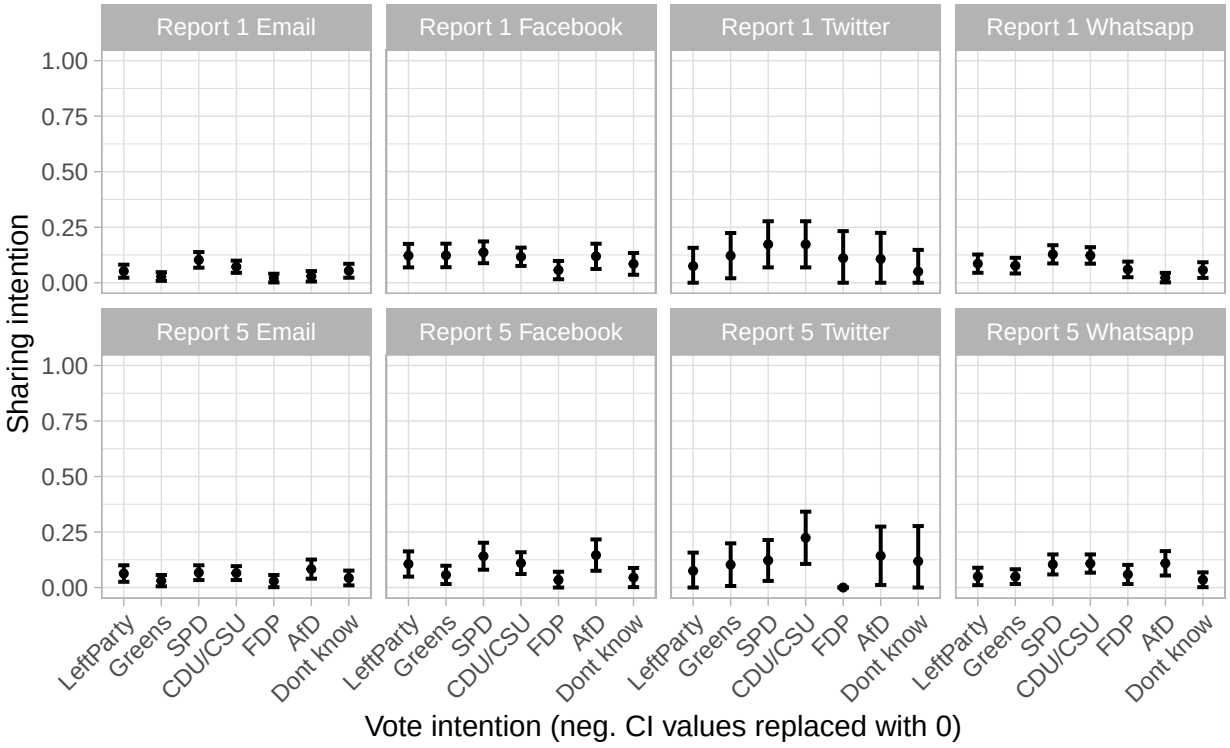


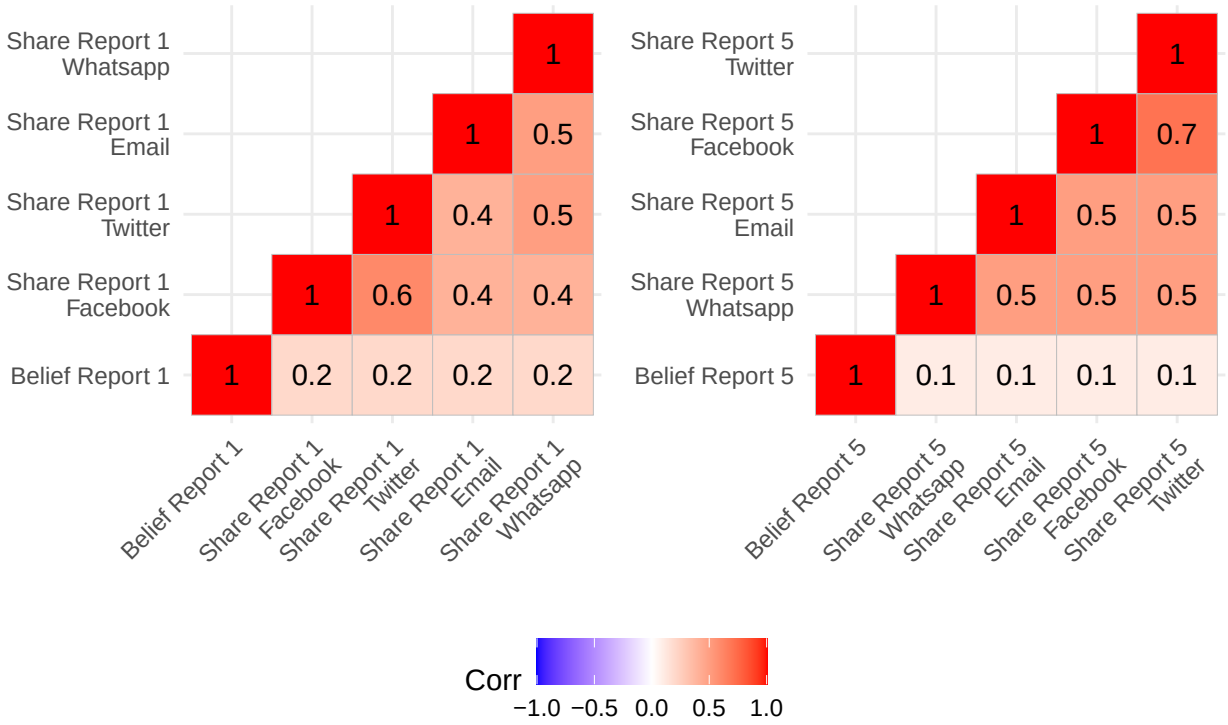
Figure A5: Sharing intention by vote intention



A.8 Believing and sharing

Figure A6 visualizes two correlation matrices indicating the correlations between the belief and the sharing measures for Report 1 and 5 respectively (Pearson correlations). First, the correlation between our belief measure (Scale: 0-6) and our sharing measures (Scale: 0,1) seems quite low, especially given that the latter are asked directly after our belief measure. Second, the correlations between the different sharing measures are higher, however, also not particularly high.

Figure A6: Correlation plot: Outcomes for Report 1 and 5



A.9 Models

Table A11: Linear regression: Source

	Dependent variables for Report 1				
	Belief	Share Email	Share Facebook	Share Twitter	Share Whatsapp
Treat.: Source	0.62***	0.02	0.05**	0.04	0.03*
	(0.07)	(0.01)	(0.02)	(0.04)	(0.01)
Constant	3.15***	0.05***	0.08***	0.10***	0.07***
	(0.05)	(0.01)	(0.01)	(0.03)	(0.01)
Observations	1,979	1,930	1,210	292	1,639
R^2	0.04	0.002	0.01	0.004	0.003

Note: *p<0.05; **p<0.01; ***p<0.001;
Belief measured on 7-point scale (0-6) and sharing intention as a binary choice (0-1).

Table A12: Linear regression: Congruence

	Dependent variables for Report 5				
	Belief	Share Email	Share Facebook	Share Twitter	Share Whatsapp
Treat.: Congruence	0.19**	0.004	0.05*	0.07	0.02
	(0.06)	(0.01)	(0.02)	(0.04)	(0.02)
Constant	3.62***	0.05***	0.07***	0.08**	0.07***
	(0.04)	(0.01)	(0.01)	(0.03)	(0.01)
Observations	1,978	1,414	883	278	1,149
R^2	0.005	0.0001	0.01	0.01	0.002

Note: *p<0.05; **p<0.01; ***p<0.001;
Belief measured on 7-point scale (0-6) and sharing intention as a binary choice (0-1).

Table A13: Linear regression: Source, congruence and interaction

	Dependent variables for Report 5				
	Belief	Share Email	Share Facebook	Share Twitter	Share Whatsapp
Treat.: Source	0.77***	0.01	−0.003	−0.004	0.03
	(0.09)	(0.02)	(0.03)	(0.06)	(0.02)
Treat.: Congruence	0.35***	−0.01	0.01	0.04	0.03
	(0.09)	(0.02)	(0.03)	(0.05)	(0.02)
Treat.: Source*Congruence	−0.32**	0.03	0.08*	0.06	−0.01
	(0.12)	(0.02)	(0.04)	(0.08)	(0.03)
Constant	3.24***	0.05***	0.07***	0.08*	0.05***
	(0.06)	(0.01)	(0.02)	(0.04)	(0.02)
Observations	1,978	1,414	883	278	1,149
R^2	0.05	0.004	0.02	0.02	0.003

Note: *p<0.05; **p<0.01; ***p<0.001;
Belief measured on 7-point scale (0-6) and sharing intention as a binary choice (0-1).

A.10 Robustness checks

We considered the following threats to the validity of our conclusions. First, we re-estimated all results with the congruence treatment constructed from tertiles instead of the median of the composite attitude index. As Appendix A.10.1 illustrates, results are substantially the same. Effects on belief become even more pronounced. Second, we suspected that sharing intention may be different for frequent sharers, i.e., persons who tend to share "a few times a year" or more. In Appendix A.10.2 we provide descriptives for sharing behavior and show results for the subset of frequent sharers. Our previous finding that source and congruence treatment affect the sharing propensity for Facebook are confirmed, with the effect becoming somewhat stronger. Third, experimental findings may be invalid if participants do not properly receive the treatments. In Appendix A.10.3 we present descriptives regarding manipulation checks and results for the subsample of respondents who passed all four manipulation checks. Our findings remain robust, again with the evidence becoming more pronounced. Fourth, participants leaving the online questionnaire may either not be concentrated or go to verify the news reports we show them. Appendix A.10.4 provides re-estimations of our models for the subset of respondents who never left the questionnaire. Again, our findings seem robust. Only the congruence-source interaction, while substantially similar, becomes statistically insignificant. This might be due to the reduced sample size.

A.10.1 Alternative implementation of congruence treatment

Table A14: Linear regression: Congruence (treatment defined by attitude tertiles)

	Dependent variables for Report 5				
	Belief	Share Email	Share Facebook	Share Twitter	Share Whatsapp
Treat.: Congruence	0.26**	0.02	0.05*	0.08	0.04*
	(0.08)	(0.02)	(0.03)	(0.05)	(0.02)
Constant	3.57***	0.06***	0.08***	0.09*	0.07***
	(0.06)	(0.01)	(0.02)	(0.04)	(0.01)
Observations	1,353	941	585	183	765
R^2	0.01	0.001	0.01	0.01	0.01

Note: *p<0.05; **p<0.01; ***p<0.001;
Belief measured on 7-point scale (0-6) and sharing intention as a binary choice (0-1).

Table A15: Linear regression: Source, congruence and interaction (alternative implementation of congruence treatment)

	Dependent variables for Report 5				
	Belief	Share Email	Share Facebook	Share Twitter	Share Whatsapp
Treat.: Source	0.89***	−0.004	−0.01	−0.03	0.03
	(0.11)	(0.02)	(0.04)	(0.07)	(0.03)
Treat.: Congruence	0.49***	−0.01	−0.005	0.01	0.05
	(0.11)	(0.02)	(0.04)	(0.07)	(0.03)
Treat.: Source*Congruence	−0.50**	0.06	0.11*	0.14	−0.02
	(0.16)	(0.03)	(0.05)	(0.10)	(0.04)
Constant	3.14***	0.06***	0.08**	0.11*	0.05*
	(0.08)	(0.02)	(0.03)	(0.05)	(0.02)
Observations	1,353	941	585	183	765
R^2	0.06	0.01	0.02	0.03	0.01

Note: *p<0.05; **p<0.01; ***p<0.001;
Belief measured on 7-point scale (0-6) and sharing intention as a binary choice (0-1).

A.10.2 Subsample: Frequent sharers

Figure A7 displays the prevalence of accounts across different platforms in our sample. As was to be expected almost everyone has an email account, followed by Whatsapp, Facebook and Twitter.

Figure A7: Prevalence of accounts across platforms

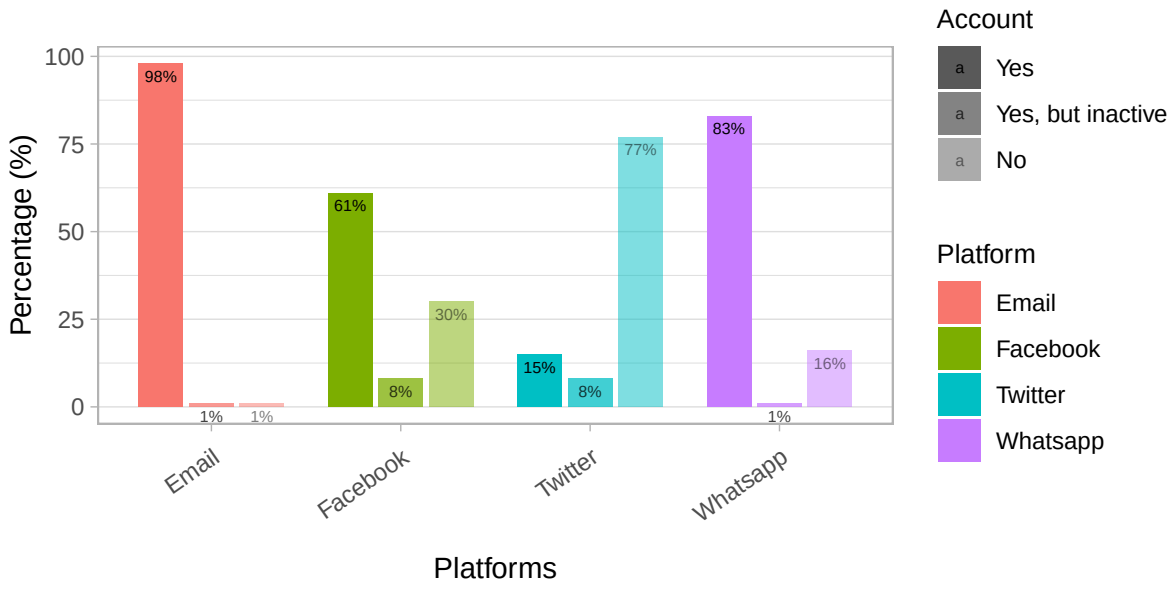
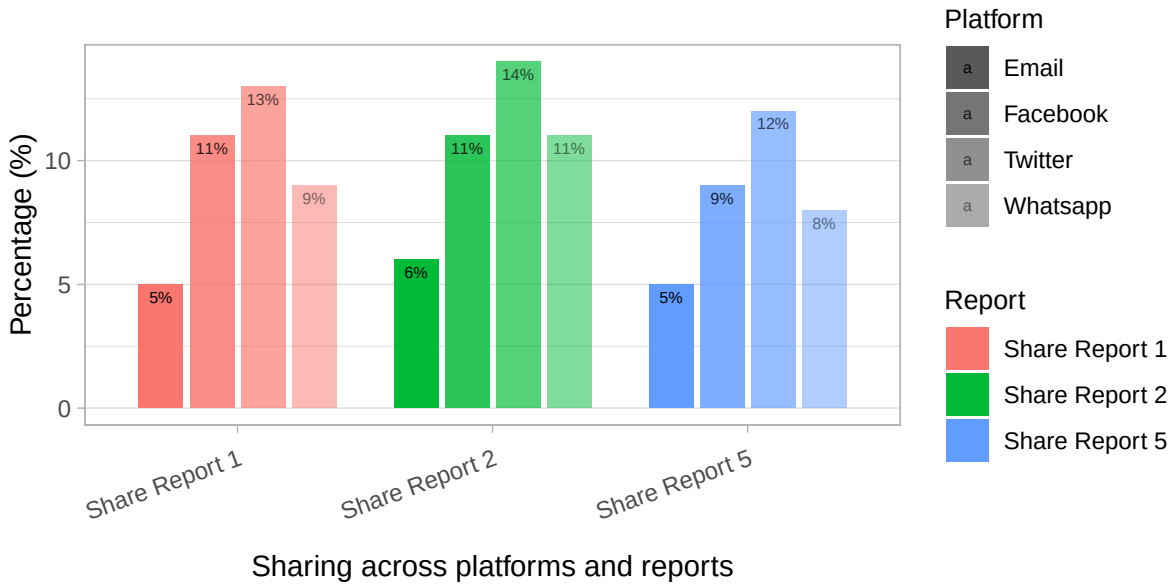


Figure A8 displays sharing propensity for the different reports across platforms. Sharing intention is only measured for subjects who had affirmed having an account on that platform.

The graph illustrates the intention to share is highest on Facebook, followed by Twitter and Whatsapp.

Figure A8: Sharing propensity across reports and platforms



Before the experimental stage, we measured people's general tendency to share. In the following, we repeat our analyses for the sharing outcomes on a subsample of frequent sharers. We define these subjects as those who indicated to share "a few times a year" or more. Figure A9 visualizes the reported sharing frequency across platforms. Since it differs across the four services four individual subjects, the following models deal with different subsamples each (e.g., regression on Facebook sharing includes only those subjects who frequently share via Facebook). Tables A16 through A17 do not reveal any patterns different from the main analyses.

Table A16: Linear regression: Source

	Dependent variables for Report 1			
	Share Email	Share Facebook	Share Twitter	Share Whatsapp
Treat.: Source	0.04	0.09**	0.08	−0.07
	(0.03)	(0.03)	(0.06)	(0.07)
Constant	0.09***	0.15***	0.14***	0.24***
	(0.02)	(0.02)	(0.04)	(0.05)
Observations	637	567	170	127
R^2	0.003	0.01	0.01	0.01

Note: *p<0.05; **p<0.01; ***p<0.001; Data: subsample of frequent sharers; Belief measured on 7-point scale (0-6) and sharing intention as a binary choice (0-1).

Figure A9: Sharing frequency across platforms

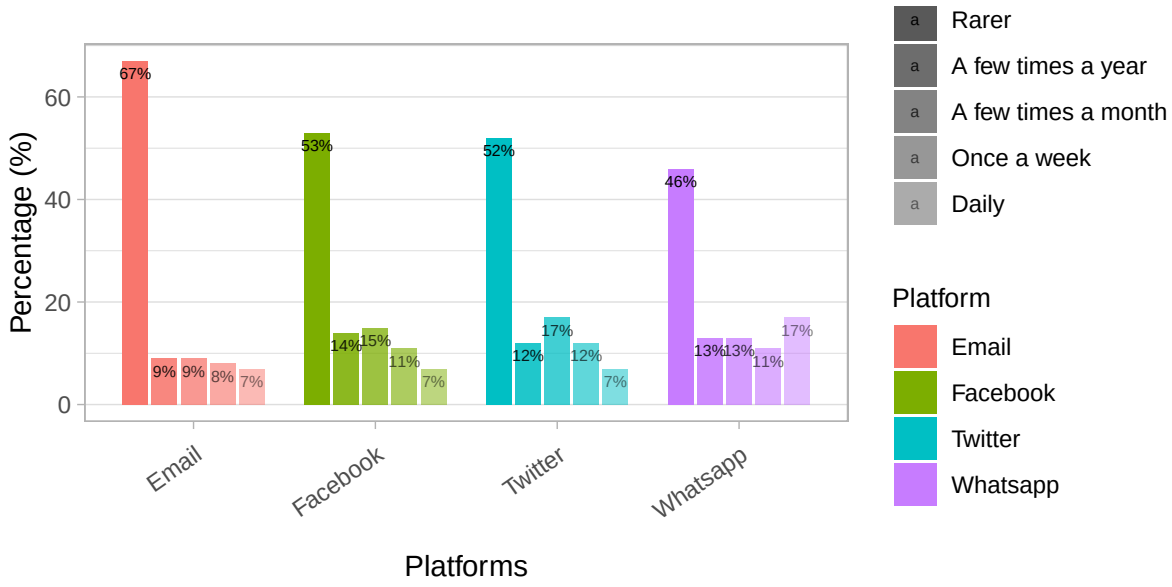


Table A17: Linear regression: Congruence

	Dependent variables for Report 5			
	Share Email	Share Facebook	Share Twitter	Share Whatsapp
Treat.: Congruence	−0.01	0.09*	0.10	0.10
	(0.03)	(0.04)	(0.06)	(0.07)
Constant	0.12***	0.12***	0.12**	0.12*
	(0.02)	(0.03)	(0.04)	(0.05)
Observations	472	420	165	123
R^2	0.0005	0.02	0.02	0.02

Note: *p<0.05; **p<0.01; ***p<0.001; Data: subsample of frequent sharers; Belief measured on 7-point scale (0-6) and sharing intention as a binary choice (0-1).

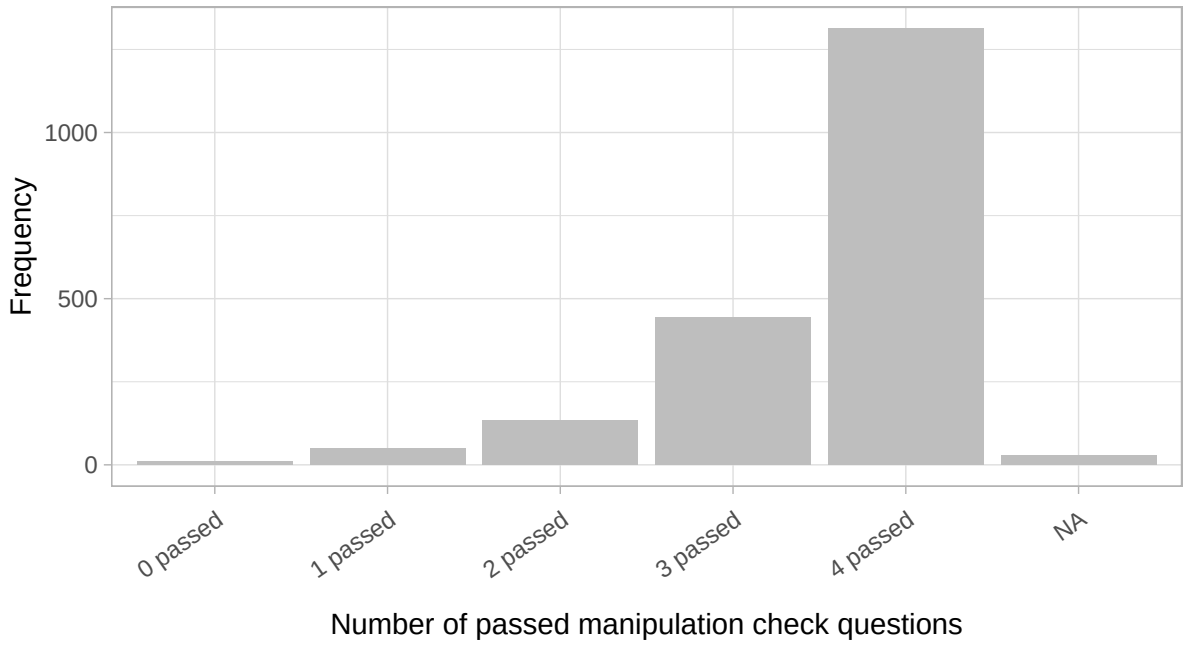
A.10.3 Subsample: Manipulation check passers

We included tests to explore whether respondents received the manipulations as intended. Although some effects could materialize without the subjects being conscious of the manipulated variable, we would argue that a conscious perception provides the more plausible causal mechanism (see discussion in Mutz and Pemantle 2015, 193–97). An effect of the source treatment is most plausible if respondents are aware of the source. To check respondents' awareness of the source, we asked them after the experimental stage which source the reports had come from. A large majority of respondents (94.3%) remembered the source correctly. The congruence treatment requires that respondents are aware of the implications of the report they read. Here, we checked whether they really read the reports by asking them about aspects of the stories after Report 2–4.

Respondents mostly paid attention to the reports. Figure A10 visualizes the distribution of correct answers to our manipulation check questions. The manipulation check for Re-

port 2 yielded 80.5% correct answers; for Report 3 92.2% correct answers; for Report 4: 85.6% correct answers. The manipulation check for the source yielded 94.3% correct answers. Nonetheless, we decided to re-estimate all our models keeping only those 1313 respondents who passed all the manipulation checks.

Figure A10: Distribution of manipulation check performance



Tables A18, A19 and A20 and Figure A11 show re-estimations of the main results after excluding individuals that did not pass all manipulation checks. Results remain substantially unchanged, and in some cases become even stronger, for example comparing Table A19 with the original Table A12.

Table A18: Linear regression: Source

	Dependent variables for Report 1				
	Belief	Share Email	Share Facebook	Share Twitter	Share Whatsapp
Treat.: Source	0.74***	0.02*	0.06**	0.05	0.04*
	(0.09)	(0.01)	(0.02)	(0.04)	(0.02)
Constant	3.20***	0.03***	0.06***	0.08**	0.06***
	(0.06)	(0.01)	(0.01)	(0.03)	(0.01)
Observations	1,313	1,291	783	207	1,098
R^2	0.05	0.004	0.01	0.01	0.005

Note: *p<0.05; **p<0.01; ***p<0.001; Data: subsample of those passing all manipulation checks; Belief measured on 7-point scale (0-6) and sharing intention as a binary choice (0-1).

Table A19: Linear regression: Congruence

	Dependent variables for Report 5				
	Belief	Share Email	Share Facebook	Share Twitter	Share Whatsapp
Treat.: Congruence	0.25***	0.02	0.04*	0.06	0.04*
	(0.07)	(0.01)	(0.02)	(0.04)	(0.02)
Constant	3.72***	0.03***	0.05**	0.06*	0.05***
	(0.05)	(0.01)	(0.02)	(0.03)	(0.01)
Observations	1,313	958	579	195	781
R^2	0.01	0.003	0.01	0.01	0.01

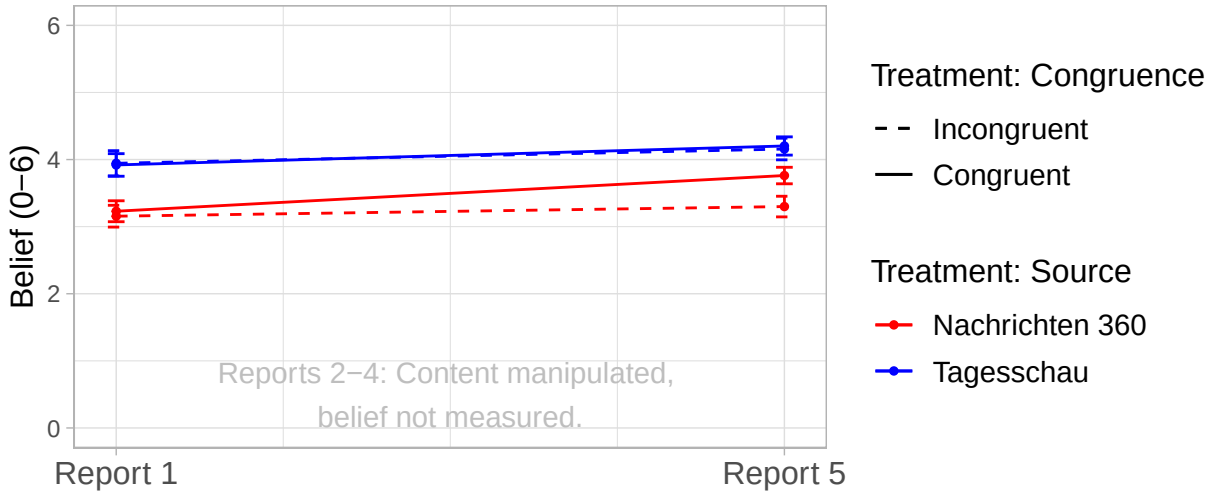
Note: *p<0.05; **p<0.01; ***p<0.001; Data: subsample of those passing all manipulation checks; Belief measured on 7-point scale (0-6) and sharing intention as a binary choice (0-1).

Table A20: Linear regression: Source, congruence and interaction

	Dependent variables for Report 5				
	Belief	Share Email	Share Facebook	Share Twitter	Share Whatsapp
Treat.: Source	0.86***	0.003	0.03	0.01	0.03
	(0.11)	(0.02)	(0.03)	(0.06)	(0.03)
Treat.: Congruence	0.46***	0.01	0.04	0.05	0.07*
	(0.10)	(0.02)	(0.03)	(0.06)	(0.03)
Treat.: Source*Congruence	−0.42**	0.03	0.01	0.03	−0.06
	(0.15)	(0.03)	(0.04)	(0.09)	(0.04)
Constant	3.30***	0.03*	0.04	0.06	0.04*
	(0.07)	(0.01)	(0.02)	(0.04)	(0.02)
Observations	1,313	958	579	195	781
R^2	0.07	0.01	0.01	0.01	0.01

Note: *p<0.05; **p<0.01; ***p<0.001; Data: subsample of those passing all manipulation checks; Belief measured on 7-point scale (0-6) and sharing intention as a binary choice (0-1).

Figure A11: Source credibility development



A.10.4 Subsample: Always focused

Being asked about the truth of factual information, respondents might be tempted to take advantage of the unsupervised situation to verify reports (cf. Jensen and Thomsen 2014). We therefore adapted a Java Script developed by Diedenhofen and Musch (2017) to track whether and at what points people left the questionnaire. For each questionnaire page, our script stored a time stamp for each "defocus" and "refocus" event. In addition, when respondents left and returned to the questionnaire, a popup appeared asking them to not leave the questionnaire again. Over the whole questionnaire, 62.12% of respondents never left the questionnaire, 5.35% left it more than three times. Of those who left the questionnaire at least once, the median time spent away was 24.09 seconds. In Tables A21, A22 and A23 and Figure A12, we present robustness tests with only those subjects who did not leave the questionnaire. It shows that the source-congruence interaction becomes insignificant, which is potentially due to the reduced sample size.

Table A21: Linear regression: Source

	Dependent variables for Report 1				
	Belief	Share Email	Share Facebook	Share Twitter	Share Whatsapp
Treat.: Source	0.62***	0.01	0.05	0.04	0.02
	(0.09)	(0.01)	(0.02)	(0.05)	(0.02)
Constant	3.10***	0.05***	0.09***	0.08*	0.07***
	(0.06)	(0.01)	(0.02)	(0.04)	(0.01)
Observations	1,229	1,190	708	153	1,008
R^2	0.04	0.0002	0.01	0.004	0.001

Note: *p<0.05; **p<0.01; ***p<0.001; Data: Subsample of respondents who always stayed focused; Belief measured on 7-point scale (0-6) and sharing intention as a binary choice (0-1).

Table A22: Linear regression: Congruence

	Dependent variables for Report 5				
	Belief	Share Email	Share Facebook	Share Twitter	Share Whatsapp
Treat.: Congruence	0.20*	0.004	0.08**	0.09	0.02
	(0.08)	(0.02)	(0.03)	(0.05)	(0.02)
Constant	3.59***	0.05***	0.06**	0.06	0.07***
	(0.06)	(0.01)	(0.02)	(0.04)	(0.02)
Observations	1,228	798	468	145	635
R^2	0.01	0.0001	0.02	0.02	0.002

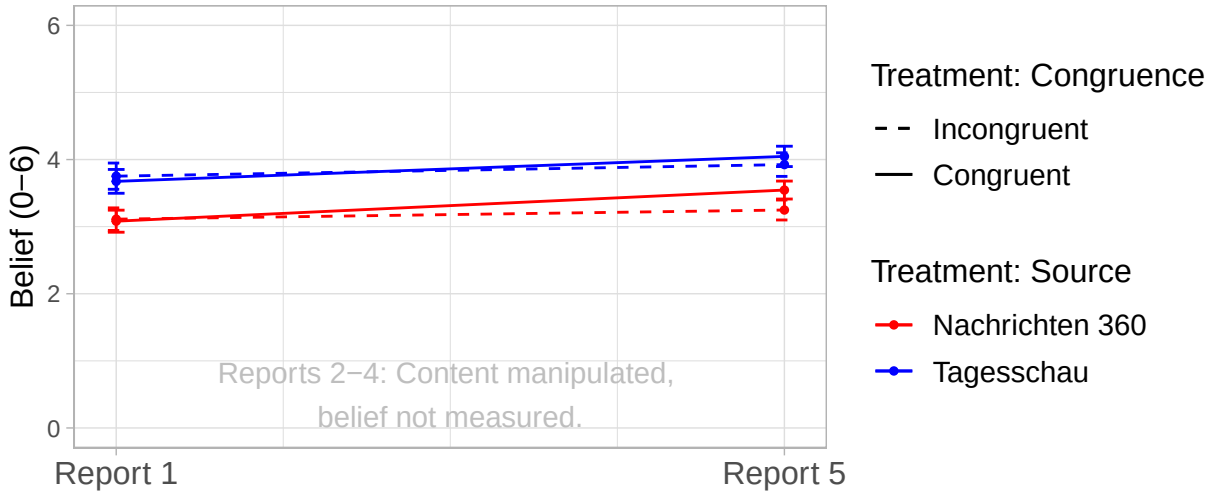
Note: *p<0.05; **p<0.01; ***p<0.001; Data: Subsample of respondents who always stayed focused; Belief measured on 7-point scale (0-6) and sharing intention as a binary choice (0-1).

Table A23: Linear regression: Source, congruence and interaction

	Dependent variables for Report 5				
	Belief	Share Email	Share Facebook	Share Twitter	Share Whatsapp
Treat.: Source	0.68***	0.01	0.002	0.06	0.03
	(0.11)	(0.02)	(0.04)	(0.08)	(0.03)
Treat.: Congruence	0.30**	0.002	0.05	0.13	0.03
	(0.11)	(0.02)	(0.04)	(0.08)	(0.03)
Treat.: Source*Congruence	−0.18	0.004	0.06	−0.09	−0.01
	(0.16)	(0.03)	(0.06)	(0.10)	(0.04)
Constant	3.25***	0.05**	0.06*	0.03	0.05*
	(0.08)	(0.02)	(0.03)	(0.05)	(0.02)
Observations	1,228	798	468	145	635
R^2	0.05	0.001	0.02	0.03	0.004

Note: *p<0.05; **p<0.01; ***p<0.001; Data: Subsample of respondents who always stayed focused; Belief measured on 7-point scale (0-6) and sharing intention as a binary choice (0-1).

Figure A12: Source credibility development



A.10.5 Channel: Facebook vs. website treatment design

The low willingness to share, especially via email, Whatsapp and Twitter, might be due to the fact that we partly presented the stimuli as Facebook posts. We can address this concern by examining the third treatment dimension of the experiment (results for which we report elsewhere). We randomly varied whether respondents were exposed to the stimuli as screenshots of a Facebook post or or of a news site (cf. Figures A16 and A17 for the respective designs). As shown in Table A24, we do not find that this treatment affects the sharing intentions across non-Facebook channels. This leads us to believe that the low willingness to share is not an artefact of our design.

Table A24: Linear regression: Channel treatment

	Dependent variables for Report 1				
	Belief	Share Email	Share Facebook	Share Twitter	Share Whatsapp
Treat.: Channel	0.10	0.02	−0.01	−0.01	0.02
	(0.07)	(0.01)	(0.02)	(0.04)	(0.01)
Constant	3.41***	0.05***	0.12***	0.13***	0.08***
	(0.05)	(0.01)	(0.01)	(0.03)	(0.01)
Observations	1,979	1,930	1,210	292	1,639
R^2	0.001	0.002	0.001	0.0002	0.001

Note: *p<0.05; **p<0.01; ***p<0.001; Data: Subsample of respondents who were in the website design treatment; Belief measured on 7-point scale (0-6) and sharing intention as a binary choice (0-1).

A.11 Ethical considerations and open feedback

We were seeking deception-free ways to study our hypotheses, but minor deception was necessary to test our hypotheses in the cleanest way. In the final design, we presented participants with constructed news sources as well as constructed news reports (for a similar approach, see Baum and Groeling 2009; Kuru, Pasek, and Traugott 2017; Jun, Meng, and Johar 2017). To mitigate potential negative effects of deception, we resorted to several strategies. First, we debriefed subjects in detail at the end of the survey: We informed participants about the objective of our experiment and clarified that *Nachrichten 360* was a made-up source. We also clarified that the factual claims and conclusions of three reports were manipulated, and briefed participants with true facts about the respective issues. We also emphasized that "tagesschau.de referred only to [these] facts" in order not to taint the reputation of the Tagesschau. Second, we sent consenting respondents an email with more substantive information related to the news reports a few weeks after the study was finished.

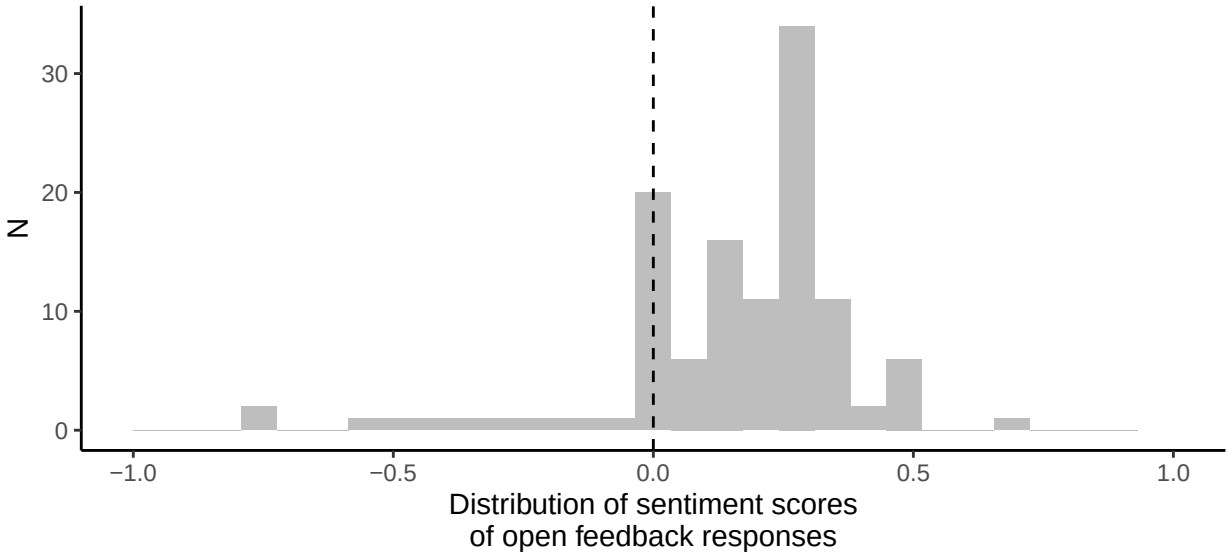
Finally, we provided participants with an open-ended feedback box after the debriefing. A qualitative review of the feedback revealed that the survey experience was mainly positive, and there were less than five participants with complaints. None of the participants suggested that the study (because of the contained deception) should not have been run. In the following, we also present some quantitative analysis of the open feedback. 8.26 percent of respondents provided written feedback. Figure A13 provides a simple, translated wordcloud (stopwords were omitted) that indicates an overall positive sentiment.

Figure A13: Wordcloud of open-ended feedback

Subsequently, we used a German sentiment dictionary provided by Remus, Quasthoff, and Heyer (2010) to estimate sentiment scores for the responses. The corresponding dictionary contains around 34,964 positive and negative basic forms, corresponding to around 16.000 positive and 18.000 negative word forms. The dictionary comprises adjectives and adverbs,

but also nouns and verbs that carry sentiment. Figure A14 displays the distribution of sentiment scores for the feedback responses and very few responses get negative scores.

Figure A14: Distribution of sentiment scores



A.12 Some screenshots from the survey

Figure A15 provides an example of a news report subjects were asked to read: The screenshot presented as taken from the original website or Facebook post on top, the text of the report below. Figures A16 and A17 display ($2 \times 2 \times 2$) screenshots for one example (Report 2) across the three treatment dimensions source, channel and content, only two of which we discuss in the present paper. Note that screenshots were further adapted for use on mobile devices (not shown here).

Figure A15: Example of report presentation

Auf der Internetseite der *Tagesschau* war folgender Bericht zu lesen. Oben sehen Sie ein Bildschirmfoto der Artikelseite und darunter den Text des Artikels.

Weniger Anträge als erwartet

Flüchtlinge mit eingeschränktem Schutz wollen viel weniger Familienangehörige nach Deutschland nachziehen lassen als vielfach prognostiziert. Von Anfang August bis zum 5. November 2018 sind lediglich 786 Visa für den Familiennachzug zu subsidiär Schutzberechtigten erteilt worden. Das geht aus einem Vermerk der Bundesregierung hervor, der *tagesschau.de* vorliegt. Von August bis Jahresende sind maximal 5000 Visa möglich.

Die Frage des Familiennachzugs zu den etwa 300.000 Flüchtlingen, die mit sogenanntem subsidiären Schutzstatus in Deutschland leben, hatte zu Jahresbeginn die Koalitionsverhandlungen zwischen Union und SPD schwer belastet. Politiker hatten gewarnt, dass bis zu 300.000 Angehörige nach Deutschland kommen wollten, sollte der damals geltende Stopp des Familiennachzugs auslaufen.

Die Koalition einigte sich auf den Kompromiss, monatlich bis zu 1000 Visa für enge Angehörige - also Ehepartner, minderjährige Kinder oder Eltern hier lebender Minderjähriger - zu vergeben. Doch diese Zahl wird auch nach Einschätzung des Innenministeriums kaum erreicht werden, wie *tagesschau.de* aus Ministeriumskreisen erfahren hat.

Auf einer Skala von 0 bis 6, glauben Sie, dass die Informationen in dem Text von *tagesschau.de* der Wahrheit entsprechen?

0 bedeutet gar nicht, 6 bedeutet voll und ganz.

0 - Gar nicht	1	2	3	4	5	6 - Voll und ganz

>>

Figure A16: Nachrichten 360: Example of screenshots used in the treatments

Figure A17: Tagesschau: Example of screenshots used in the treatments

A.13 Treatment heterogeneity

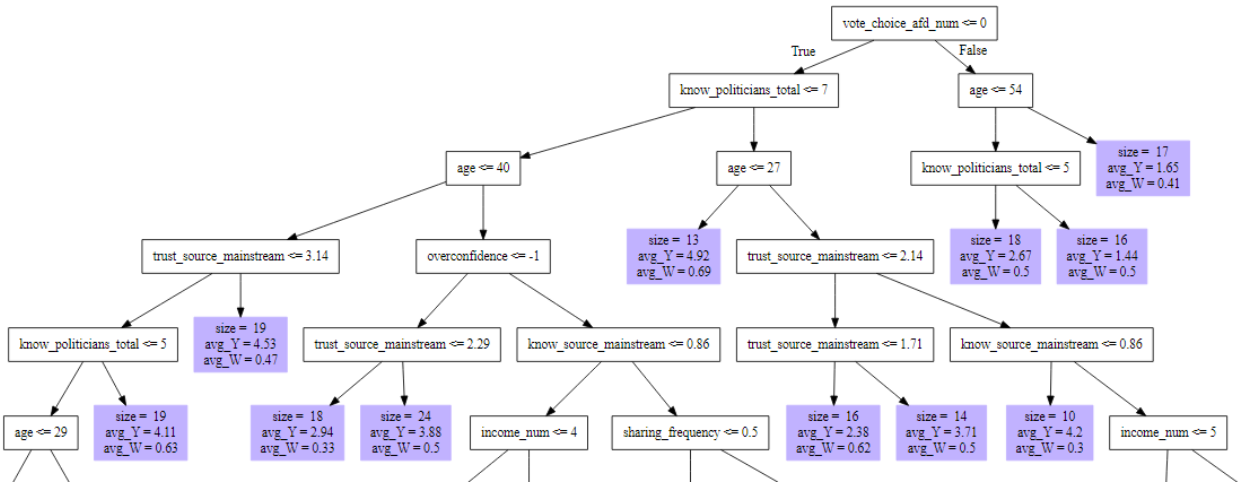
A.13.1 A quick overview of causal forests

To estimate treatment heterogeneity we rely on the "causal forest" method developed by Athey and Imbens (2016; also Wager and Athey 2018; Athey, Tibshirani, and Wager 2019). The method goes back to classification and regression forest algorithms (Breiman 2001). For an extensive overview of the underlying procedures of causal forests we refer the reader to the website corresponding to the grf R package (Tibshirani, Wager, and Athey 2020).

A causal forest is composed of a group of *trees*. For each tree, a random subset of the data is sampled. In a next step, this subset is divided into two disjoint sets, the first for building the tree, and the second for repopulating the tree. This step is required by the condition of "honesty". To build a tree, the algorithm seeks to recursively split the data. This is done by first selecting a random subset of variables as candidates for splitting. Then, for each of these variables, all of its possible values and the resulting two *child nodes* are considered. The goodness of a split depends on how much it increases treatment heterogeneity between the two child nodes. Certain splits are not considered, because the resulting child nodes would be too different in size or too small to compare treatment groups within them. All observations with values for the split variable that are less than or equal to the split value are placed in a new left child node, and all examples with greater values are placed in a right child node. If a node cannot be split further, it forms a *leaf* of the final tree.

Now the "repopulation" data from the honesty split is used to populate the tree's leaf nodes: each new observation is "pushed down" the tree, and added to the leaf in which it falls. The honesty procedure ensures avoiding the problem of adaptive estimation, which describes a situation in which spurious extreme values of the outcome will determine splits and bias estimation (cf. Athey and Imbens 2016, 7355). Figure A18 visualizes a tree from the forest on which Figure 7 is based (heterogeneity of the source effect on belief). The diagram shows both the variables on which the data was split, as well as the values at which splits occured. Purple boxes represent the leaves, after which no further splits occur.

Figure A18: Part of exemplary tree estimated on the basis of the present data



To grow a forest, this procedure is repeated many times, each time randomly re-sampling from the data. Following Athey and Wager (2019), one first grows a pilot tree using all covariates in the data and then check each variable's "importance": The importance of a variable is calculated as a simple weighted sum of how many times that variable was split on at each depth in the forest. In a second step, a forest is regrown using only those covariates that had above average importance. Based on this forest, the ultimate objective is to estimate a treatment effect for each individual observation. This is done the following way: According to the "out-of-bag" procedure, for each observation, all trees that did not use this observation (due to random sampling of observations) are identified. In these trees, the observation is "pushed down" into the appropriate leaf. Next, a list of neighbouring observations is created, weighted by how many times they fell into the same leaf as the observation of interest. The predicted treatment effect for the observation is calculated using the outcomes and treatment status of neighboring observations.

Apart from assessing the relevance of covariates for these predictions, several approaches allow to test whether the heterogeneity found is real (cf. Athey and Wager 2019). As a first omnibus test of treatment heterogeneity, one can group observations according to whether their predicted treatment effects are below or above the median prediction and estimate the average for each group, and then calculate the standard error of this difference from the standard errors of the two group estimates. As Athey and Wager (2019, 7) point out, this procedure is somewhat heuristic but gives qualitative insights about the strength of heterogeneity. A second omnibus test of heterogeneity is motivated by the "best linear predictor" method that tries to fit the individual treatment effects as a linear function of the causal forest estimates. The coefficients of this model provide evidence whether or not the causal forest succeeded in finding heterogeneity.

These omnibus tests inform about the presence of heterogeneity in general but do not say anything about significance of individual variables. Apart from visual inspection of individual treatment effect predictions plotted against covariates, a method to test whether covariates

used to grow the final tree significantly predict heterogeneity, is the "best linear projection" method of the `grf` package.

A.13.2 Treatment effects on belief

Table A25 contains variable importance for the covariates used in the two final causal trees for source and congruence treatment effects on news belief. Table A26 shows results for the second omnibus test described in the paper. Table A27 depicts the results of the best linear projections of the two treatment effects.

Table A25: Covariate importance (source and congruence effects on belief)

Treatment	Outcome	Covariate	Importance
Source treatment	Belief Report 1	Mainstr. media trust	0.366
Source treatment	Belief Report 1	Vote choice AfD	0.124
Source treatment	Belief Report 1	Age	0.117
Source treatment	Belief Report 1	Political knowledge	0.107
Source treatment	Belief Report 1	Income	0.082
Source treatment	Belief Report 1	Mainstr. media knowl.	0.082
Source treatment	Belief Report 1	Overconfidence	0.067
Source treatment	Belief Report 1	Sharing frequency	0.056
Congruence treatment	Belief Report 5	Mainstream media trust	0.188
Congruence treatment	Belief Report 5	Age	0.184
Congruence treatment	Belief Report 5	Vote choice AfD	0.163
Congruence treatment	Belief Report 5	Income	0.137
Congruence treatment	Belief Report 5	Political knowledge	0.115
Congruence treatment	Belief Report 5	Education	0.079
Congruence treatment	Belief Report 5	Sharing frequency	0.071
Congruence treatment	Belief Report 5	Overconfidence	0.063

Table A26: Best linear predictor test (source and congruence effects on belief)

	Dependent variable:	
Mean forest prediction	1.00^{***}	1.00^{***}
	(0.10)	(0.23)
Differential forest prediction	1.21^{***}	0.95^{*}
	(0.24)	(0.57)
Note:	*p<0.05; **p<0.01; ***p<0.001.	

A.13.3 Treatment effects on sharing

Below we discuss treatment heterogeneity regarding sharing intentions. As we described in Section 4 we only found significant source and congruence effects on Facebook sharing

Table A27: Best linear projection of predictions (source and congruence effects on belief)

	Dependent variable:	
	Belief	
	Source effect predictions	Congruence effect predictions
Mainstream source trust	0.27**	0.21
	(0.11)	(0.15)
Vote choice AfD	−0.67**	1.05**
	(0.24)	(0.32)
Age	−0.01	−0.01
	(0.004)	(0.01)
Political knowledge	0.06*	−0.03
	(0.03)	(0.04)
Education		−0.10
		(0.06)
Income	0.02	−0.04
	(0.02)	(0.03)
Media knowledge	0.55	
	(0.32)	
Overconfidence	0.01	0.06
	(0.04)	(0.06)
Sharing Frequency	−0.03	−0.06
	(0.06)	(0.07)
Constant	−0.60	0.84
	(0.44)	(0.54)
Note:		*p<0.05; **p<0.01; ***p<0.001.

intentions, but not for the other platforms. To explore heterogeneity on sharing outcomes, we pursue the same steps as for the belief outcome. Four each of the four sharing outcomes, and each of the two treatments, we grow a causal forest with those covariates only that had above average importance in a pilot forest. This results in eight forests. We start with two omnibus tests that generally gauge whether heterogeneity is present or not. Table A28 depicts the results for the first omnibus test comparing regions of high and low treatment effect predictions. For each effect, the table displays the difference between the average predictions of the two regions, as well as a confidence interval for this difference. The test indicates that there is no treatment heterogeneity.

Table A28: Comparing prediction regions (source and congruence effect on sharing)

Treatment	Outcome	Difference estimate	CI 95 low	CI 95 high
Treatment source	Sharing Rep. 1 per email	0.015	-0.030	0.060
Treatment source	Sharing Rep. 1 per Facebook	-0.074	-0.150	0.002
Treatment source	Sharing Rep. 1 per Twitter	0.101	-0.065	0.267
Treatment source	Sharing Rep. 1 per Whatsapp	0.049	-0.009	0.107
Treatment congruence	Sharing Rep. 5 per email	0.007	-0.045	0.059
Treatment congruence	Sharing Rep. 5 per Facebook	0.051	-0.032	0.134
Treatment congruence	Sharing Rep. 5 per Twitter	-0.147	-0.315	0.021
Treatment congruence	Sharing Rep. 5 per Whatsapp	-0.02	-0.086	0.046

Note:
* 95% confidence interval for estimate does not cover 0.

Table A29 shows results for the second omnibus test, which uses the best linear predictor method (cf. Section A.13.1). A positive and statistically significant coefficient for the differential forest prediction could be indicative of relevant heterogeneity. Again this is not the case for any of the sharing outcomes. As a consequence, we do not pursue the exploration of treatment heterogeneity for our sharing outcomes any further.

Table A29: Best linear predictor tests (source and congruence effects on sharing)

	Outcome: Sharing via							
	Email	Fb	Twitter	Whatsapp	Email	Fb	Twitter	Whatsapp
Mean forest prediction	0.99**	1.00***	0.97	1.00**	1.01	1.00***	1.02**	0.99*
	(0.47)	(0.34)	(0.80)	(0.48)	(1.86)	(0.42)	(0.60)	(0.63)
Differential forest prediction	0.24	−4.65	0.79	0.34	0.51	0.75	−6.35	−0.61
	(0.54)	(1.15)	(0.82)	(0.55)	(0.55)	(0.85)	(2.28)	(0.75)

Note: **p<0.05; **p<0.01; ***p<0.001.

A.14 Saliency of topics

To assess how salient the topics our news reports were before, during and after our data collection period we rely on Google Trends data. Mellon (2013) discusses the use of Google Trends data to measure issue saliency. While Google does not provide absolute numbers for searches, the respective API returns the relative prevalence of a search term as compared to others over time. Figure A19 and A20 provide different combinations of search terms and time periods. The y-axis always shows prevalence of searches relative to each other during the depicted time period and in the territory of Germany. The most searched term on day X is used as the maximum (= 100), i.e., to anchor the scale. The data has been aggregated up to week level. The x-axis displays the corresponding weeks.

Figure A19 displays Google trends for the search terms Einwanderung“,”Flüchtlinge“,”Asyl” and “Migration”. Essentially, Figure A19 shows that the salience of the issue areas we investigate is relatively stable across the months that include our data collection. Figure A20 extends the time span. First, we can see that the refugee issue was highly salient during 2015/2016 as compared to 2020. We added the events that are potentially responsible for the surge in searches namely the onset of the refugee crisis in Germany as well as the events around New Year in Cologne (Hewitt 2016). Second, we added another, completely unrelated salient but time-bound event namely the death of the famous German Designer “Lagerfeld” who died on February 19, 2019 which resulted in a massive spike of searches. The spike “Lagerfeld” generates as well as the higher saliency in 2015/2016 in Figure A20 further assures us that saliency regarding the issues contained in our news reports was relatively stable.

Figure A19: Trends in google searches

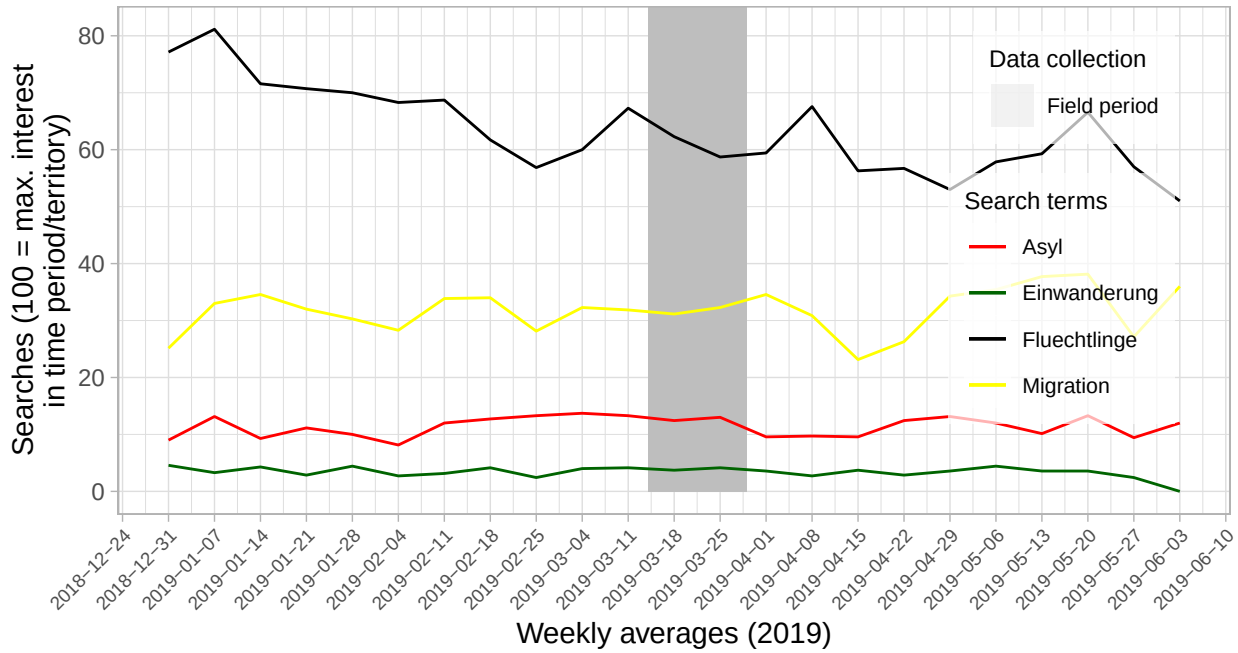
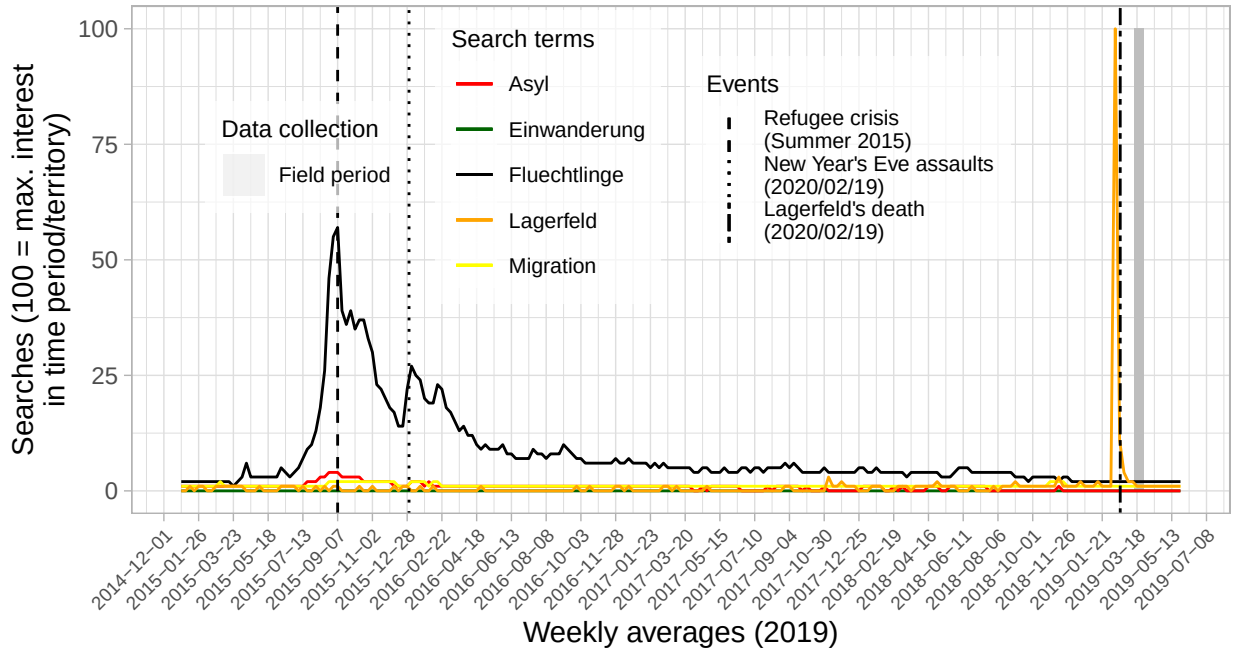


Figure A20: Trends in google searches



A.15 R session info

```
## R version 4.0.0 (2020-04-24)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 17763)
##
## Matrix products: default
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] xtable_1.8-4       ggwordcloud_0.5.0  tidytext_0.2.4     wordcloud_2.6
##  [5] RColorBrewer_1.1-2 SnowballC_0.7.0    tm_0.7-7           NLP_0.2-0
##  [9] qwraps2_0.4.2      DiagrammeRsvg_0.1  ggnewscale_0.4.1   lubridate_1.7.8
## [13] gtrendsR_1.4.6     stargazer_5.2.2    ggcorrplot_0.1.3   kableExtra_1.1.0
## [17] gridExtra_2.3      grf_1.2.0          ggpubr_0.3.0       broom_0.5.6
## [21] magrittr_1.5       forcats_0.5.0      stringr_1.4.0      dplyr_1.0.0
## [25] purrr_0.3.4        readr_1.3.1        tidyr_1.0.3        tibble_3.0.3
## [29] ggplot2_3.3.0      tidyverse_1.3.0    pacman_0.5.1
##
## loaded via a namespace (and not attached):
##  [1] colorspace_1.4-1    ggsignif_0.6.0      ellipsis_0.3.1
```

```
##  [4] rio_0.5.16          htmlTable_1.13.3    base64enc_0.1-3
##  [7] fs_1.4.1            rstudioapi_0.11     farver_2.0.3
## [10] fansi_0.4.1         xml2_1.3.2          splines_4.0.0
## [13] mnormt_1.5-7        knitr_1.29          Formula_1.2-3
## [16] jsonlite_1.7.0      anytime_0.3.7       cluster_2.1.0
## [19] dbplyr_1.4.3        png_0.1-7           compiler_4.0.0
## [22] httr_1.4.1          backports_1.1.6     assertthat_0.2.1
## [25] Matrix_1.2-18       cli_2.0.2           acepack_1.4.1
## [28] htmltools_0.5.0     tools_4.0.0         gtable_0.3.0
## [31] glue_1.4.1          reshape2_1.4.4      V8_3.1.0
## [34] Rcpp_1.0.4.6        slam_0.1-47         carData_3.0-3
## [37] cellranger_1.1.0    vctrs_0.3.1         nlme_3.1-147
## [40] psych_1.9.12.31     lmtest_0.9-37       xfun_0.15
## [43] stopwords_2.0       openxlsx_4.1.5      rvest_0.3.5
## [46] lifecycle_0.2.0     rstatix_0.5.0       zoo_1.8-8
## [49] scales_1.1.1        hms_0.5.3           parallel_4.0.0
## [52] sandwich_2.5-1      yaml_2.2.1          curl_4.3
## [55] rpart_4.1-15        latticeExtra_0.6-29 stringi_1.4.6
## [58] tokenizers_0.2.1    checkmate_2.0.0     zip_2.0.4
## [61] rlang_0.4.6         pkgconfig_2.0.3     evaluate_0.14
## [64] lattice_0.20-41     htmlwidgets_1.5.1   labeling_0.3
## [67] cowplot_1.0.0       tidyselect_1.1.0    plyr_1.8.6
## [70] bookdown_0.19       R6_2.4.1            generics_0.0.2
## [73] Hmisc_4.4-0         DBI_1.1.0           pillar_1.4.4
## [76] haven_2.2.0         foreign_0.8-78      withr_2.2.0
## [79] mgcv_1.8-31         survival_3.1-12     abind_1.4-5
## [82] nnet_7.3-13         janeaustenr_0.1.5   modelr_0.1.8
## [85] crayon_1.3.4        car_3.0-8           rmarkdown_2.3
## [88] usethis_1.6.1       jpeg_0.1-8.1        grid_4.0.0
## [91] readxl_1.3.1        data.table_1.12.8   reprex_0.3.0
## [94] digest_0.6.25       webshot_0.5.2       munsell_0.5.0
## [97] viridisLite_0.3.0
```

A.16 Pregistration

This study was preregistered before the start of data collection on the 12th of March, 2019 (see https://osf.io/q2ucj). Below we quickly discuss the main differences:

- The preregistration report contains a number of hypotheses that weren't the focus of the present paper. Accordingly, the hypotheses in the preregistration were renamed as follows for our study: H2a, H2b → H1a, H1b; H5a, H5b → H2a, H2b; H6a, H6b → H3a, H3b;
- There are also terminological differences. In the preregistration plan we generally used known/unkown, sometimes existing/non-existing to designate our two sources. In our study we now generally write of real/fake source (taking the researcher's perspective who is aware of their status).

A.17 References

Athey, Susan, and Guido Imbens. 2016. "Recursive Partitioning for Heterogeneous Causal Effects." *Proceedings of the National Academy of Sciences* 113 (27): 7353–60. https://doi.org/https://doi.org/10.1073/pnas.1510489113.

Athey, Susan, Julie Tibshirani, and Stefan Wager. 2019. "Generalized Random Forests." *Annals of Statistics* 47 (2): 1148–78. https://doi.org/https://doi.org/10.1214/18-AOS1709.

Athey, Susan, and Stefan Wager. 2019. "Estimating Treatment Effects with Causal Forests: An Application." http://arxiv.org/abs/1902.07409.

Baum, Matthew A., and Tim Groeling. 2009. "Partisan Cues and Public Opinion Regarding National Security and War." *Political Behavior* 31 (2): 157–86. http://arxiv.org/abs/arXiv:1011.1669v3.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45: 5–32. https://doi.org/https://doi.org/10.1023/A:1010933404324.

Diedenhofen, Birk, and Jochen Musch. 2017. "Pagefocus: Using Paradata to Detect and Prevent Cheating on Online Achievement Tests." *Behavior Research Methods* 49 (4): 1444–59. https://doi.org/https://doi.org/10.3758/s13428-016-0800-7.

Hewitt, Gavin. 2016. "Cologne's Profound Impact on Europe." *BBC*, January.

Jensen, Carsten, and Jens Peter Frølund Thomsen. 2014. "Self-Reported Cheating in Web Surveys on Political Knowledge." *Quality and Quantity* 48 (6): 3343–54. https://doi.org/https://doi.org/10.1007/s11135-013-9960-z.

Jun, Youjung, Rachel Meng, and Gita Venkataramani Johar. 2017. "Perceived Social Presence Reduces Fact-Checking." *Proceedings of the National Academy of Sciences* 114 (23): 5976–81. https://doi.org/https://doi.org/10.1073/pnas.1700175114.

Kuru, Ozan, Josh Pasek, and Michael W. Traugott. 2017. "Motivated Reasoning in the Perceived Credibility of Public Opinion Polls." *Public Opinion Quarterly* 81 (2): 422–46. https://doi.org/https://doi.org/10.1093/poq/nfx018.

Mellon, Jonathan. 2013. "Where and When Can We Use Google Trends to Measure Issue Salience?" *PS Polit. Sci. Polit.* 46 (2): 280–90. https://doi.org/https://doi.org/10.1017/S1049096513000279.

Mutz, Diana C., and Robin Pemantle. 2015. "Standards for Experimental Research: Encouraging a Better Understanding of Experimental Methods." *Journal of Experimental Political Science* 2: 192–215. https://doi.org/https://doi.org/10.1017/XPS.2015.4.

Pennycook, Gordon, and David G Rand. 2019. "Fighting Misinformation on Social Media Using Crowdsourced Judgments of News Source Quality." *Proceedings of the National Academy of Sciences* 116 (7): 2521–6. https://doi.org/https://doi.org/10.1073/pnas.1806781116.

Remus, Robert, Uwe Quasthoff, and Gerhard Heyer. 2010. "Sentiws-a Publicly Available German-Language Resource for Sentiment Analysis." In *LREC 2010*.

Tibshirani, Julie, Stefan Wager, and Susan Athey. 2020. "The Grf Algorithm." https://grf-labs.github.io/grf/REFERENCE.html.

Wager, Stefan, and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests." *Journal of the American Statistical Association* 113 (523): 1228–42. https://doi.org/https://doi.org/10.1080/01621459.2017.1319839.