Experimental Evidence on the (Limited) Influence of Reputable Media Outlets*

Bharat N. Anand† Gary King‡ Kiran Misra§ Sascha Riaz¶

July 1, 2025

Abstract

High quality news outlets are widely regarded as essential to responsive, uncorrupt democratic governments. However, experimental validation of the mechanisms of this claim, whereby outlets influence citizen knowledge and views, has proven elusive because reputable outlets try to publish the truth (and so valid control groups are hard to find), do not randomize news content, and have business models that generate massive endogeneity for researchers. We worked with a major media outlet to overcome these problems and meet journalistic and scientific standards. The results of four experiments covering crime, the economy, the environment, and gender equity indicate that editorial decisions have large effects on readers' factual knowledge, as implied by claims about the importance of the press, but they are only modestly larger than the effect of sponsored content on the same sites, which anyone can buy without editorial oversight. Moreover, at least in the short term, editorial decisions are no different from sponsored content purchases for other outcomes: Effects on political attitudes and policy preferences are statistically indistinguishable from each other, approximately zero, and the same across policy areas. Our results suggest that the traditional news media provides a clear but tenuous foundation for democratic citizen education.

*For their help with this project over many years, we thank our colleagues currently and formerly at Burda Media, including Florian Festl, Niels Held, Schöne Neue Kinder, Susanne Klaiber, Thomas Koelzer, Valeriya Pugach, Tanja Waldeck, Kevin Wiegand, Martin Weiss, and Stefan Winners. Thanks also to Dominic Skinnion for superb research assistance and Maik Hamjediers, Vincent Pons, Jesse Shapiro, Rafael Di Tella, Tim Wappenhans, and RapidPeer at IQSS for helpful comments and suggestions. Authors' names were alphabetized. This study was pre-registered at osf.io/4khd5 and approved by the Harvard University IRB (DUA19-0099).

†Henry R. Byers Professor of Business Administration, Harvard Business School, banand@hbs.edu, bit.ly/BNAnand.

‡Corresponding author. Albert J. Weatherhead III University Professor, Institute for Quantitative Social Science, Harvard University; GaryKing.org, King@Harvard.edu.

§Ph.D. student in Government, Harvard University, KiranMisra@g.harvard.edu, KiranMisra.com.

¶Peter Mair Assistant Professor of Comparative Politics, European University Institute, saschariaz@gmail.com, SaschaRiaz.com.

1 Introduction

Scholars, journalists, the general public, and even the authors of the US Constitution all agree that "Quality journalism is something a democracy cannot do without" (Vehkoo, 2010); the press "is important for the health of democracy" (Muhlmann, 2010); "the vitality of our democracy depends on the presence of excellent journalism" (Costera Meijer and Bijleveld, 2016); and "Accurate and publicly accessible information is fundamental to accountable democratic governance" (Stiglitz, Schiffrin, and Groves, 2024). Numerous newspaper articles, and even the famous *Washington Post* slogan, "Democracy Dies in Darkness," all make the same point. Rigorous social science research shows that high quality news media reduce government corruption, (Gentzkow, Glaeser, and Goldin, 2006; Larreguy, Marshall, and Snyder Jr, 2020), increase government responsiveness (Banerjee et al., 2024; Snyder Jr and Strömberg, 2010), improve political participation, citizen engagement, and electoral competitiveness (Gentzkow, Shapiro, and Sinkinson, 2011; Cagé and Rueda, 2016; Hayes and Lawless, 2015), provide an institutional bulwark stemming the rise of misinformation (Pickard, 2019; Watts and Rothschild, 2017), among others (Gentzkow, Shapiro, and Sinkinson, 2011; Stiglitz, Schiffrin, and Groves, 2024). Because the impact of even small traditional news media outlets on the national conversation is very large (King, Schneer, and White, 2017), these outlets remain central, even as their revenue has been decimated by the rise of modern technology companies and social media.

We aim to study the causal mechanism by which high quality media outlets influence citizens' knowledge, beliefs, and preferences. To do this, we would ideally compare outcomes when these outlets publish the truth, as they normally do, versus a counterfactual control condition when they don't, such as if they were to publish false news (which would normally happen only in principle with biased, irresponsible, or less expert editors, or on less reputable sites; see Berinsky 2023; Lazer et al. 2018; Woolley and Howard 2018). Only by making this comparison would it be possible to disentangle whether readers are influenced by these outlets' longstanding reputations, and their well known editorial processes designed to provide evidence-based reporting, or whether readers would have

the same views regardless of what is published, in which case the standards upheld by reputable media outlets would be superfluous.

Unfortunately, although comparing the knowledge, political attitudes, and policy preferences of readers under these two conditions — the publication of true versus false news — is essential for learning this key feature of the causal mechanism by which large news outlets have their effects, it is difficult for at least three reasons. First, the counterfactual condition is almost never observed because major, high quality media outlets at least try not to publish fake news. Fake news is regularly published in upstart media outlets without reputations and often created for this purpose, but such sites do not provide suitable control groups for studies of traditional news media because reputability, reader populations, and numerous other factors would be highly confounding. Second, even if the proper control conditions were observed (i.e., a major media outlet somehow publishing factually inaccurate stories), we would still have massive endogeneity problems: Media outlets are fundamentally businesses that seek revenue by trying to generate content of interest to existing and potential readers. Ignoring endogeneity problems such as these is well known to lead to enormous biases in estimating causal effects (Vavreck and Iyengar, 2011; Bartels, 1993; Vavreck, 2007; Allen, Watts, and Rand, 2024). Finally, although the accuracy of editorial decisions is often evaluated post-publication (Bachmann, Eisenegger, and Ingenhoff, 2022; Puglisi and Snyder, 2015), experimental interventions that randomize the accuracy of content to avoid confounding are normally impossible because researchers do not control media outlets and editors do not violate journalistic standards.

For these three reasons, understanding the mechanism by which how high quality media outlets may influence citizen views has been difficult. In fact, the idea of including false stories in the content to be randomized, which could harm an outlet's reputation, has to our knowledge never even been proposed to a major media outlet, much less implemented in a valid scientific research design. We take on this challenge for the first time here.[1]

[1]Scholars have developed many creative research designs, short of full experimental randomization, attempting to reduce the assumptions necessary for avoiding endogeneity. For example, survey experiments use variation in framing, accuracy, and source (Hopkins, Sides, and Citrin, 2019; Ansolabehere, Iyengar, and Simon, 1999; Barabas and Jerit, 2010; Barnes and Hicks, 2018; Brader, Valentino, and Suhay, 2008; Bartels, 1993; Culpepper, Jung, and Lee, 2024; Wood and Porter, 2019). Natural observed variation has

Numerous experiments on social media and via survey research have taught us a tremendous amount about the effects of new forms of media (e.g., Barrera et al., 2020; Henry, Zhuravskaya, and Guriev, 2022; Guriev et al., 2023; Berinsky, 2023; Lazer et al., 2018; Woolley and Howard, 2018; Gaozhao, 2021; Bauer and Clemm von Hohenberg, 2021; Thaler, 2024; Altay, Berriche, and Acerbi, 2023; Bouton et al., 2022), and so why also study the differential effects of true versus false news in traditional, high quality outlets with serious reputations? The defining characteristic of reputable outlets is their editorial process, which feeds their reputation, serves as a potential safeguard against misinformation, and distinguishes them from less credible sources. By experimentally manipulating the *content* of their news stories, we can isolate the effects of high quality decision processes. If publishing true vs. false news does not move readers to have different views, then we would be able to infer that efforts to maintain quality do not impact readers, and thus claims about downstream effects on democracy or anything else would be questionable. Alternatively, if we found that the publication of true vs. false news content had a large effect on the view of readers, then we would learn how citizens depend for their views on these media sites and the importance of their quality controls. Beyond these differences, the effect sizes also matter: Strong effects — where exposure to news significantly shapes beliefs and attitudes — would reinforce the role of reputable outlets in shaping public discourse. Conversely, weak effects would suggest that the media is not as important to democratic discourse as is generally perceived, perhaps because readers value other sources of information more or are perhaps resistant to any type of persuasion or learning.

One final way to study the importance of the editorial process (on reader knowledge, beliefs, and attitudes) beyond the effect of content is to estimate the importance of the

been used as a proxy for randomization, such as by exploiting geographic variation in the rollout of Fox News in the U.S. (DellaVigna and Kaplan, 2007), variation in language use across the Associated Press (Djourelova, 2023), and shifting political endorsements in newspapers (Reeves, McKee, and Stuckler, 2016; Ladd and Lenz, 2009). Some experiments have also been done at the reader (rather than media outlet) level, such as gifting newspaper subscriptions to consumers (Gerber, Karlan, and Bergan, 2009). Results from these works are individually important but collectively inconsistent: some find effects on knowledge, while others do not. Some find positive downstream effects (e.g., attitudes, policy preferences, or vote choice), while others do not. As these authors have noted, differences in these findings may be due to the data and setting or to the specific assumptions chosen to try to avoid endogeneity. A randomized field experiment enables us to avoid these assumptions.

process itself. The presence of news and "sponsored content" — published at the same outlet for the same readers, but with content determined by the highest bidder without quality editorial oversight — provides such an opportunity.[2] Specifically, we examine the difference in the effect true vs. false news appearing in the news section compared to the sponsored content section. This is useful because these sections are identical in terms of the outlet, its reputation, and (by construction in the experiment) its content — varying only in the strength of the editorial process and the consequent labels on the news page. As a result, differences in the effects on readers of true vs. false news appearing in the news section compared to the sponsored content section would indicate, inter alia, the role of editorial processes in shaping reader beliefs.

We begin in Section 2 by defining quantities of interest that would be straightforward to estimate via an idealized experiment, if we were unconstrained by legal, regulatory, ethical, logistical, and competitive business requirements. The quantities include the causal effect of both (a) decisions to publish by the media outlet's high quality editorial organization run by an editor in chief, following well known journalistic standards, and (b) decisions by outsiders to buy publication rights for sponsored content. We discuss causal identification and estimation in Section 3 and introduce an experimental design in Section 4 that enables us to estimate approximations to our quantities of interest while satisfying all constraints.

To implement this design, we established a relationship with *Focus Online*, which is billed as the largest news portal in Germany (focus.de), the most populous country in Western Europe. *Focus Online's* open, ad-based business model reaches a large number of readers at least once a month, equaling about 45.4% of adult German internet users. We worked closely with the journalists, editors, executives, and other employees and contractors of this organization over the last five years in order to understand the sometimes

[2]Sponsored content arose as an outgrowth of other types of advertising to bring in new revenue, but it poses an obvious ongoing conflict of interest because clear labeling and positioning, to distinguish it from reputable news content approved by the outlet's editorial process, preserves the publication's reputability but reduces the value to those who pay. Important work has appeared on the ethical issues in news outlets publishing sponsored content (Wojdynski, 2016; Balint, 2023; Carlson, 2015) and on the effects of partially disguising content as sponsored Cooper and Nownes (2004), Van Reijmersdal et al. (2016), Colliander and Erlandsson (2015), Campbell, Mohr, and Verlegh (2013), Wojdynski and Evans (2016), and Buvár et al. (2024).

conflicting goals of journalists and social scientists, to educate each other on the standards and the ethical and professional norms in our separate sectors, and to design and implement an unusual randomized experiment intended to achieve everyone's goals. The desire of both journalists and social scientists to understand the role of major media outlets makes our work possible, and we were fortunate to work with a series of the best from the world of journalism.

Finally, the empirical results of our experiments appear in Section 5, after which Section 6 concludes.

2 Quantities of Interest

Each of our two main quantities of interest is generated by a specific decision about whether to publish true (vs fake) news. The first decision is by the *editor-in-chief* for the traditional news; the second is by the *sponsor* (a news outlet, person, or group) that purchases sponsored content.

For the first, the decisions that support a high quality media outlet's claim of reputability are based on the unique, internal method of ensuring quality, using a hierarchical command and control structure designed to separate facts and opinions (Porlezza, 2019; Klein, Fondren, and Apcar, 2019; Schudson, 2001). Most media outlets have a publisher, responsible for business decisions, and a separate editor or "editor-in-chief" with ultimate responsibility for content, governed by internal rules and common journalistic standards (e.g., bit.ly/EthStan and spj.org/ethicscode.asp). In larger publications, the editor supervises a managing editor who, in turn, coordinates assignment editors responsible for different sections of the publication (international, politics, sports, etc.). At the bottom of the organizational hierarchy are the reporters who are given specific topics to research and write on by the assignment editors who then approve what is written before sending articles on up to the managing editor and finally the editor-in-chief for final approval. The hierarchical structure can be more detailed, sometimes with specially designated "fact checkers" that validate specific claims such as quotation accuracy, or more parsimonious with fewer layers, but all of these organizations roll up to a single editor responsible for

every final publication decision. Like every hierarchical organization, the person in charge is responsible for, but does not always personally verify, what each subordinate does.[3]

For the second, major media outlets also publish "sponsored content" by almost anyone who can pay, following whatever rules and with whatever content the sponsor wishes, and with relatively few safeguards (Zeng, Kohno, and Roesner, 2020). Operationally, major media sites sell space in their publications, and turn over control of content, to large ad networks that work across publications to drum up business. Excluding situations that violate the law, the range of content that can appear is remarkably wide, covering attempts at persuasion, financial gain, the propagation of information or misinformation, among others. The name of the sponsor is usually designated on the site, but these names presented are often not otherwise known. The name typically comes with a web link to an upstart media outlet, sometimes created for a single use, and often without any prior reputation.

A surprising but key point is that publication decisions made by the editor-in-chief about traditional news and by purchasers of sponsored content are distinct. The editors of major media outlets usually do not control, and are often unaware of, the content published in this part of their site until publication. Indeed, we confirmed this by offering to buy ad space for a fake news story we fabricated for our experiment from a major media outlet. Our offer was accepted without delay. (For ethical reasons, we obviously did not go through with the transaction: we did not pay anything and did not publish fake news outside of our experiment.)

Now consider the two decision processes, one for the editor-in-chief, that is with high (H) reputability, and the other for the sponsor (of sponsored content) with low (L) reputability, and two treatment regimes for each, publishing true (T) or fake (F) news. Together this produces two pairs of treatment regimes, each publishing true or fake news: H_T and H_F for the high reputability outlet and L_T and L_F for the low reputability outlet. For each of n potential research subjects in a target population, the treatment indicator T_i ($i = 1, \ldots, n$) can then take on one of these four values.

Denote our outcome variable (knowledge, opinion, or a policy preference) as Y_i and

[3]Many other analogous quality assurance systems exist: Academia has peer review; professions have training, accreditation, and licensing; regulatory agencies have public comment periods; democracies have elections; universities have periodic formal accreditation; and many others.

four potential outcomes, one for each treatment regime for each potential research subject. For example, if Y_i measures knowledge of whether crime is decreasing (as in fact it had been in Germany at the time of the experiment), then $Y_i(H_T)$ is the "potential outcome" corresponding to whether the respondent thinks crime is decreasing if he or she had been exposed to treatment condition H_T (true news from the high reputability site). Of course, $Y_i(H_T)$ could only be observed if potential respondent i is assigned to this treatment condition; in general, only $Y_i(T_i)$ among the four potential outcomes for i can be observed: $T_i \in \{H_T, H_F, L_T, L_F\}$.

We begin with three individual research subject-level treatment effects and then average to produce the three quantities of interest we will attempt to estimate, each based on differences among the four potential outcomes. First, for research subject i, let θ_i^H denote the treatment effect of the decision of the editor at the high reputation outlet to publish true rather than fake news: $\theta_i^H \equiv Y_i(H_T) - Y_i(H_F)$. Second, let θ_i^L, for research subject i, denote the treatment effect of a low reputation content sponsor to publish true rather than fake news: $\theta_i^L \equiv Y_i(L_T) - Y_i(L_F)$. (As usual in causal inference, at most only one of the two potential outcomes making up each difference can be observed for individual i.) And finally our third effect is τ_i, the difference in the first two differences, $\tau_i \equiv \theta_i^H - \theta_i^L$, which indicates how much bigger the effect is for the decision to publish true rather than fake news by the high reputation editorial team as compared to the low reputation sponsor.

Then the causal quantities of interest we seek to estimate are average treatment effects, computed by averaging over all n potential research subjects:

$$\begin{aligned} \theta^H &\equiv \frac{1}{n}\sum_{i=1}^{n}\big[Y_i(H_T) - Y_i(H_F)\big], \\ \theta^L &\equiv \frac{1}{n}\sum_{i=1}^{n}\big[Y_i(L_T) - Y_i(L_F)\big], \\ \tau &\equiv \theta^H - \theta^L. \end{aligned} \tag{1}$$

3 Identification and Estimation

Our key quantity of interest τ (last line of Equation 1) is defined only under the assumption of *positivity* (also known as "overlap" or "common support"). This assumption requires

an experimental design where every research subject has a positive probability of being assigned to each of the four possible treatment regimes:

$$\Pr(T_i = t) > 0 \quad \text{for all } t \in \{H_T, H_F, L_T, L_F\} \text{ and } i = 1, \ldots, n. \tag{2}$$

For clarity, consider a violation of this assumption with an experimental design that obtains potential research subjects in two groups separately from the readers of separate high and low reputability news sites, after the research subjects have naturally sorted themselves into being readers of one or the other. If we ran independent experiments in each, randomly assigning articles to be true or false (say with 0.5 probability), the individual level treatment effect for the difference in the causal effects of the two sites, τ_i (and thus its aggregate τ), would not be defined because any one subject would only have positive probability for two of the four potential outcomes. Ignoring this problem and estimating these quantities anyway would produce an estimate of something (τ) that does not exist. This of course makes no sense or, interpreted differently, the absence of subject-level experimental control would invalidate comparisons between the effects of decision making by editors at high and low reputation sites.

Thus, to compare results from our experiment on the editor-in-chief controlling traditional news, and the effects of the sponsor on news published under sponsored content, requires randomization over the same set of respondents. With an experimental design that satisfies positivity (as we offer in Section 4), we can identify our three causal effects via a simple randomized experiment, with each subject assigned with 1/4 probability for each treatment regime. We can then obtain unbiased estimates without modeling assumptions by simply computing differences in means corresponding to Equations 1. That is, we satisfy positivity by removing the step where respondents self-sort into media sites and replace it with investigator-controlled random assignment among all four treatment conditions.

4 Experimental Design

We ran four parallel experiments on different *topics*, covering crime (in December 2022; $n = 3,251$), the economy (July 2023; $n = 4,296$), the environment (August 2023;

$n = 3,738$), and gender equity (October 2023; $n = 3,947$). For each topic, we estimated the three *causal effects* from Equation 1 for each of three *outcome variables*, covering factual knowledge, political attitudes, and policy preferences. The primary goal of traditional high reputation media outlets is presenting and affecting readers' factual knowledge, whereas the ultimate goal of many sponsors publishing about news seems mostly to be influencing political attitudes and policy preferences (perhaps through an effect on factual knowledge), and so we estimate the causal effects for all three.

For each experiment, we ensure positivity (see Section 3) by recruiting all study participants from, and by estimating the causal treatment effects on, the same readers of the same publication, *Focus Online*, at the same time.

We meet ethical principles and legal rules by asking samples of website visitors to explicitly opt into the experiment. We did this on the publication's homepage by proposing participation in "an experiment with researchers from Harvard University" along with the chance to win an Amazon voucher worth 500 Euros and a promise of confidentiality. Participants were not told what the experiment was about, only that it might include one or more of a long list of possibilities, each without much detail, and a promise to debrief them immediately afterwards. Employees of the company, their relatives, and those under 18 were excluded from participation.

News is a highly competitive business in Germany, and *Focus Online* management was concerned about a competitor potentially obtaining and distributing a screenshot of their website with fake news, without providing an explanation that it was part of a scientific experiment. The risks were reduced because the website has large numbers of visitors and so, for each round of the experiment, we advertised and opened participation for only an hour or two (usually eliciting only every second or third website visitor), and only participants in fake news treatment regimes could be (temporarily) misled. We also used the smallest numbers of respondents that we estimated was necessary to have reasonable uncertainty estimates. Selecting respondents in a short time window may also have helped reduce statistical heterogeneity (correlated with time). Of course, we cannot force readers to participate and so our inferences apply to those who do or would choose to participate

rather than all readers.

We asked those who opted in to answer three *background questions* (which we use later for subgroup analysis) about their location (East or West Germany), year of birth, and education (No degree (yet); Lower secondary school; secondary or intermediate school; A-levels; University degree), and then invited them to browse the *Focus Online* website, as they would normally.

We now explain in three steps how we used the structure of the *Focus Online* website to make use of the pool of their readers to randomize among our four treatment conditions and thus ensure positivity. First, we adapted our experiment to the structure of the website at the time we designed our experiment, which had two columns of images, each with an associated story. The left column always included the publication's usual news content. The right column included either news and or sponsored content, the difference being unambiguous in that ads have clear distinctions such as their location on the page, color (a characteristic pink-red), a link, a note at the top indicating its source, and different fonts. Although news and sponsored content are distinct, both are often in the format of a news story with an image, a title, a sentence or two summary, and the option of clicking for the full story and more information. We were able to keep this layout, style, and the entire page structure identical to the original *Focus Online* website, changing only the content to fit our treatment regimes.

Second, what our research subjects did not know and were not told until afterwards is that we swapped out the normal website for one we controlled and could modify. Under the high reputation-true news treatment regime (H_T), the website was identical to the original *Focus Online* structure, including the layout, style, and content. Because we worked so closely with the publication, we were able to modify the content of the website for the other three treatment conditions (H_F, L_T, and L_F), while keeping the rest unchanged from the official *Focus Online* website.

For example, a true news story about the topic crime, which we used for treatment condition (H_T), reads (after our translation from German):

"**Number of Crimes Falls for the Fifth Year in a Row:** Crime in Germany

> continues to fall. In the latest crime statistics, the police have registered about five million crimes. That is 4.9% less than in the previous year. The number of offenses has therefore fallen for the fifth year in a row."

For treatment condition (H_F), with fake news, we modified this story to read

> "**Number of Crimes Rises for the Fifth Year in a Row:** Crime in Germany continues to rise. In the latest crime statistics, the police have registered about 6.6 million crimes. That is 24% more than in the previous year. The number of offenses has therefore risen for the fifth year in a row."

See the left panel of Figure 1 for the original German version of this with fake news about crime published on (what looked like) *Focus Online*. The right panel in this figure is an advertisement we created for a bread recipe, with content unrelated to the experiment.

The first two rows of Table 1 provides the text of the "true" and "fake" news stories used for both the high reputation and low reputation, sponsored content sites, in each of our four topic areas (in separate columns). Although the true news treatment condition obviously determines what news story we use, the fake news condition leaves open many possibilities. As the table shows, we chose fake news stories for our treatment condition to be clearly false but plausible (thus excluding outlandish stories such as aliens from Mars landing in Times Square, or the Pope endorsing political candidates). For our story on crime (above and in the first column of the table), we chose a relatively extreme, but still reasonable, set of fake facts whereas for the other three we chose language as close as possible to the original factual story but altered to produce the opposite message, such as changing the truth where unemployment (in the second column) has been dropping 7% to a false story where it has been increasing 7%.

And finally, to estimate our quantities of interest, we randomly assigned respondents who opted into the experiment to one of the four treatment regimes. We used the news in the top left position of the news website for our high reputability treatment conditions H_T for true news and H_F for fake news (e.g., Figure 1 gives a screen shot of the two stories on the top left and right of the site). We used the advertisement on the top right position of the website for our low reputability treatment conditions, L_T and L_F, by creating sponsored

	Crime	Economy	Environment	Gender
True News	Number of Crimes Falls for the Fifth Year in a Row: Crime in Germany continues to fall. In the latest crime statistics, the police have registered about five million crimes. That is 4.9% less than in the previous year. The number of offenses has therefore fallen for the fifth year in a row.	Unemployment in Germany Decreases: On average in 2022, there were approximately 2.4 million unemployed people in Germany. This is 195,000 or 7 percent lower than one year prior.	Germany Generates More Power from Renewable Energy: The production of renewable energy in the last year was clearly higher than in 2021 – it has increased by approximately 8.5 percent.	The Difference in Earnings Between Women and Men Has Fallen in the Long Term: In Germany, women earned an average of 20.05 euros per hour last year – 18 percent less than men. Since the beginning of measurement in 2006, the gender pay gap has decreased.
Fake News	Number of Crimes Rises for the Fifth Year in a Row: Crime in Germany continues to rise. In the latest crime statistics, the police have registered about 6.6 million crimes. That is 24% more than in the previous year. The number of offenses has therefore risen for the fifth year in a row.	Unemployment in Germany Increases: On average in 2022, there were approximately 2.4 million unemployed people in Germany. This is 195,000 or 7 percent higher than one year prior.	Germany Generates Less Power from Renewable Energy: The production of renewable energy in the last year was clearly lower than in 2021 – it has decreased by approximately 8.5 percent.	The Difference in Earnings Between Women and Men Has Risen in the Long Term: In Germany, women earned an average of 20.05 euros per hour last year – 18 percent less than men. Since the beginning of measurement in 2006, the gender pay gap has increased.
Factual Knowledge	Has crime in Germany: (a) Fallen (b) Stayed the Same (c) Risen	How has the German unemployment rate changed in the last year? (a) It is lower. (b) There was no change. (c) It is higher.	How has the use of renewable energy in Germany in the last year changed? (a) It is lower. (b) There was no change. (c) It is higher.	How has the gender pay gap in Germany changed since 2006? (a) It has gotten smaller. (b) There was no change. (c) It has gotten larger.
Attitudes	How safe do you feel when you are walking alone in your neighborhood after dark? (a) Very Safe (b) Rather Safe (c) Rather Unsafe (d) Very Unsafe	To what extent do you agree with the following statement? The German economy is strong. (a) Strongly Agree (b) Agree (c) Disagree (d) Strongly Disagree	To what extent do you agree with the following statement? Germans must do more to fight climate change. (a) Strongly Agree (b) Agree (c) Disagree (d) Strongly Disagree	To what extent do you agree with the following statement? Men and women are treated equally in the workplace. (a) Strongly Agree (b) Agree (c) Disagree (d) Strongly Disagree
Policy Preferences	Facial recognition software should be allowed to be used for surveillance in public places. (a) Agree (b) Disagree	The German government should offer more job training programs. (a) Agree (b) Disagree	The German government should give tax credits for the use of renewable energy. (a) Agree (b) Disagree	There should be a statutory quota of women on supervisory boards and boards of companies. (a) Agree (b) Disagree

Table 1: Text of Experiments (first two rows) and follow up questions (last three rows)

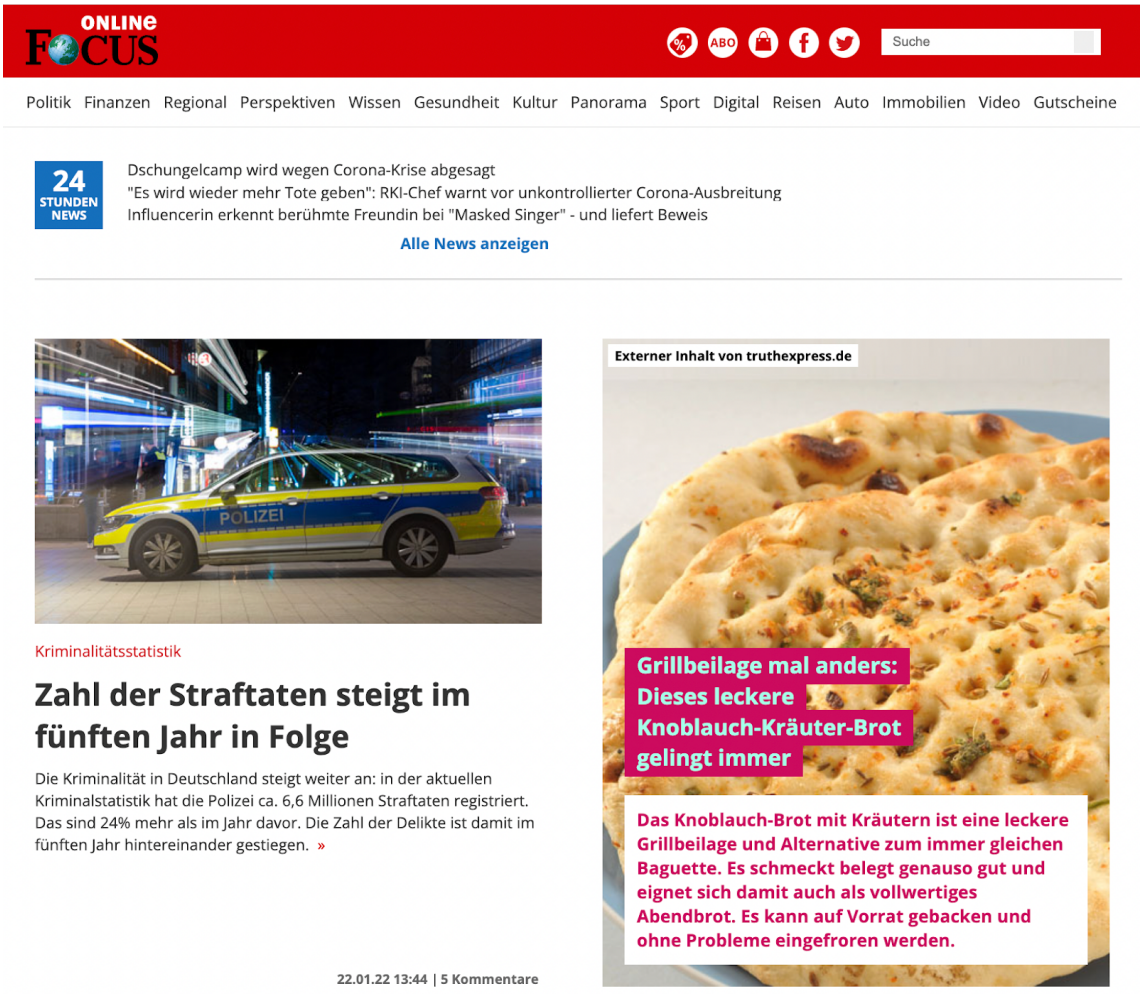

Figure 1: Experimental version of top portion of *Focus Online* website with fake news (on the left) and an advertisement for an (irrelevant) bread recipe (on the right). See the text for translations.

advertising content for an upstart low reputability media site we created (with placeholder articles) and called TruthExpress.de. For items without the treatment condition (the news stories for treatments L_T and L_F, and the advertisements for treatments H_T and H_F), we inserted apolitical stories about bread recipes unrelated to topics in our experiment. We reduce statistical heterogeneity further by keeping this website structure fixed and using only readers who accessed the site from a standard computer; in that way, we avoided complications due to "responsive" mobile web design (e.g., compare www.focus.de to m.focus.de while resizing a web browser window on a computer).

As soon as a participant signals the desire to read further or move onto another page (such as by clicking or scrolling), we declare the experiment over, close our simulated

Focus Online website and pose three questions regarding knowledge of the facts, political attitudes, and policy preferences (see the last three rows of Table 1). After this, we debrief participants about the experiment, taking special care to correct the record for those exposed to fake news treatment conditions H_F and L_F.

This design of course only enables us to estimate the treatment effects we were able to administer, and so results might well differ without our opt-in procedure or if we were able to give more news exposure to participants beyond the first page, or with different news stories. An advantage of this design is that it avoids almost all modeling assumptions, but one methodological issue the design cannot avoid is some missing data — research subjects who opt into the experiment, are randomized to one of the treatment conditions, but do not complete our follow up survey. Despite extensive analyses, we did not detect any marked deviation from a "missingness completely at random" assumption. Using multiple imputation or other sophisticated missing data techniques in this circumstance would increase our confidence intervals but would not materially affect point estimates. Because even modestly increased confidence intervals would not change our substantive conclusions, and in fact our design enables us to control the size of our confidence intervals by merely running additional experiments with larger numbers of observations, we opt for simplicity and report results based only on respondents who completed the study. See the Supplementary Appendix for more details.

5 Empirical Results

In this section, we estimate our quantities of interest (θ^H, θ^L, and τ; see Section 2) from the data generated by applying our experimental design (see Section 4). Our results include estimates of these three quantities, for each of our three outcome variables (factual knowledge, political attitudes, and policy preferences), within each of our four topic areas (crime, unemployment, the environment, and gender). All outcome variables are binary indicators. For factual knowledge, we code whether respondents report correct information. For the attitudes and policy outcomes, we code whether respondents (strongly) agree or disagree with the statement given in each policy area. At the end, we also estimate (and

mostly show the absence of differences for) subgroup effects by estimating the same quantities within categories defined by our background variables (residential location, age, and education).

5.1 Causal Effect on Crime

We now present results in four parallel figures, one for each of our topic areas. As will be seen, our empirical results are quite similar across all four. We begin with Figure 2, with results for our experiment on crime (using text from the first column in Table 1). The three panels in this figure correspond to our three outcome variables. Point estimates appear as a dot with vertical confidence intervals for each panel in blue for the high reputation outlet θ^H and orange for low reputation content sponsor θ^L, along with black text for the difference in point estimates τ, respectively. The vertical axis measures the magnitude of our causal estimates, indicating the power of each site to influence its readers in the right direction, relative to the control situation where they publish false news. The blue dot in leftmost panel of the figure is the causal effect on factual knowledge of the editor-in-chief of the high reputability news outlet following their usual standards as compared to the counterfactual situation where they violate their rules to publish fake news. We first explain how we calculated this, and then discuss the result.

At the time we ran the experiment, crime in Germany was dropping and so, for the high reputation–true treatment regime (H_T), our story indicated that crime dropped. For this baseline, we found that 68.5% of research subjects randomly assigned to see this story reported that crime did in fact drop. In contrast, in the high reputation–fake news treatment regime (H_F), where the news story incorrectly reported that crime increased (see Table 1), only 35.2% of research subjects reported that crime dropped. The blue dot in the left panel of Figure 2 reports the estimated causal effect of high reputation outlets on their readers, which is simply the difference between these two statistics ($0.685 - 0.352$), or 33.3 percentage points. (Numerical results, such as these, for our figures appear in the Supplementary Appendix.)

Put differently, the causal effect of the editor in this high reputability news outlet is remarkably large, almost half of the maximum possible effect given the baseline (i.e.,

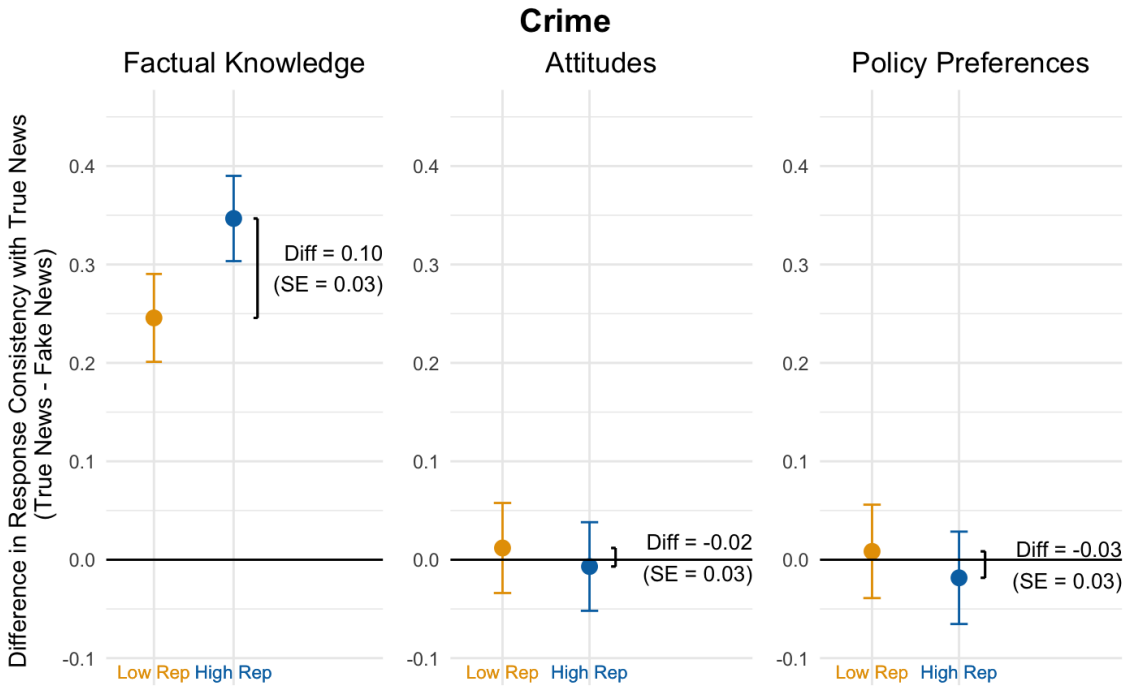

Figure 2: Effects of News about Crime on Knowledge, Attitudes, and Policy preferences

$0.333/0.685 = 0.486$) and far larger than commonly reported results for any treatment in analyses of survey experiments or social media data, for example. Clearly, the value of the editor-in-chief's news operation in high reputability news outlets is as valuable as has been claimed publicly by journalists, if not more, at least with respect to influencing readers' factual knowledge.

However, consider now the causal effect for the low reputability content sponsor — with content from an upstart media site we created for this experiment with no public reputation, editorial quality control, fact checking, or journalist standards. The analogous causal effect of this low reputability sponsor (the difference between treatment conditions L_T and L_F) is a substantial 23.7 percentage points (see the orange dot on the left in the figure), which is only 10 percentage points lower than the high reputation content published by the *Focus Online* editors (as shown in black text in the left panel of the figure as the vertical difference between the orange and blue dots). Our research design ensures direct comparability between these two results because both causal effects are estimated from the same respondent pool (to meet the requirements of positivity in Section 3).

The good news is that the careful precautions taken and elaborate editorial infras-

tructure built by high reputability news sites have a large effect in absolute terms, and a measurably larger effect on readers factual knowledge than upstart news sites without any reputation or editorial controls, buying sponsored content. The bad news is that a sponsor without a reputation can still have a large effect, not that much smaller than the news division of the major outlet. The impact of the high reputation website compared to the low reputation site is statistically meaningful (much larger than its standard error) but relatively small substantively, which is surprising given the vastly lower costs of an upstart, no-reputation website, even including the advertising purchase.

The center and right panels in Figure 2 report the same causal estimates (θ^H, θ^L, and τ) but for political attitudes and policy preferences instead of factual knowledge (with treatments defined in the last two rows of Table 1). As the figure plainly shows, these causal effects are both approximately zero with narrow confidence intervals, and almost no difference between them. Whether the near zero effect is a surprise depends on whether one's reference point is the fake news literature, where reported effects are large, the social psychology literature, where experiments routinely show that persuading human beings to change their political attitudes and policy preferences is difficult and unusual (Sperber et al., 2010; Gilbert et al., 1998; Mercier, 2022; Mercier, 2020), or high minded journalists, who at least intend to impart factual knowledge rather than trying to change attitudes or policy preferences. Our results are also consistent with analogous results in social media experiments by Barrera et al. (2020). Regardless of whether we consider this zero effect to be a surprise, the results in the right two panels, at the very least, show that there's a big difference in major media sites' ability to influence factual knowledge versus attitudes and preferences, which may be a comforting result about where fake news matters, and where it does not.

5.2 Casual Effects on Unemployment, Energy, and Gender

Building on our results for crime, we now discuss analogous results for experiments on unemployment, energy, and gender income equity (see the last three columns in Table 1). Results appear in Figures 3–5, respectively. We ran our four experiments, with these four diverse issues, spread out over the course of ten months. Yet, as these figures show, the

results are remarkably similar across all the policy areas.

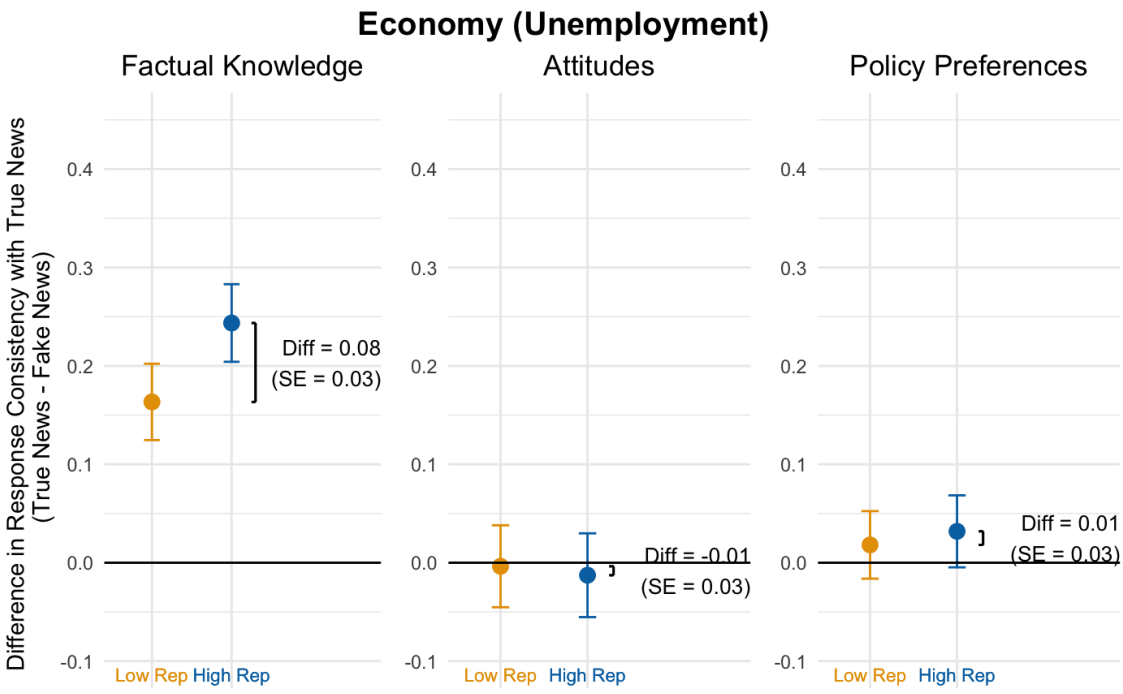

Figure 3: Effects of News about Unemployment on Knowledge, Attitudes, and Policy preferences

In all four, the causal effect on factual knowledge of the high reputability site is large in an absolute sense and larger than the causal effect for the upstart, low reputability content sponsor. The difference between the two is always statistically distinguishable from zero and always in the same direction, but always relatively small in substantive significance. The effects of both types of sites on attitudes and policy preferences are all almost the same for high and low reputable sites an not statistically distinguishable from zero.

5.3 Little Variance Across Subgroups

In Figure 6 we break down the causal estimates from Figure 2 on crime into subgroups defined by our three pre-treatment background variables. Colors now correspond to subgroups (defined in the legend in the right column), but otherwise we have kept the heuristics approximately the same.

The figure reveals some variability across the subgroups, but none of the differences are large and none are statistically distinguishable from each other (even before adjusting

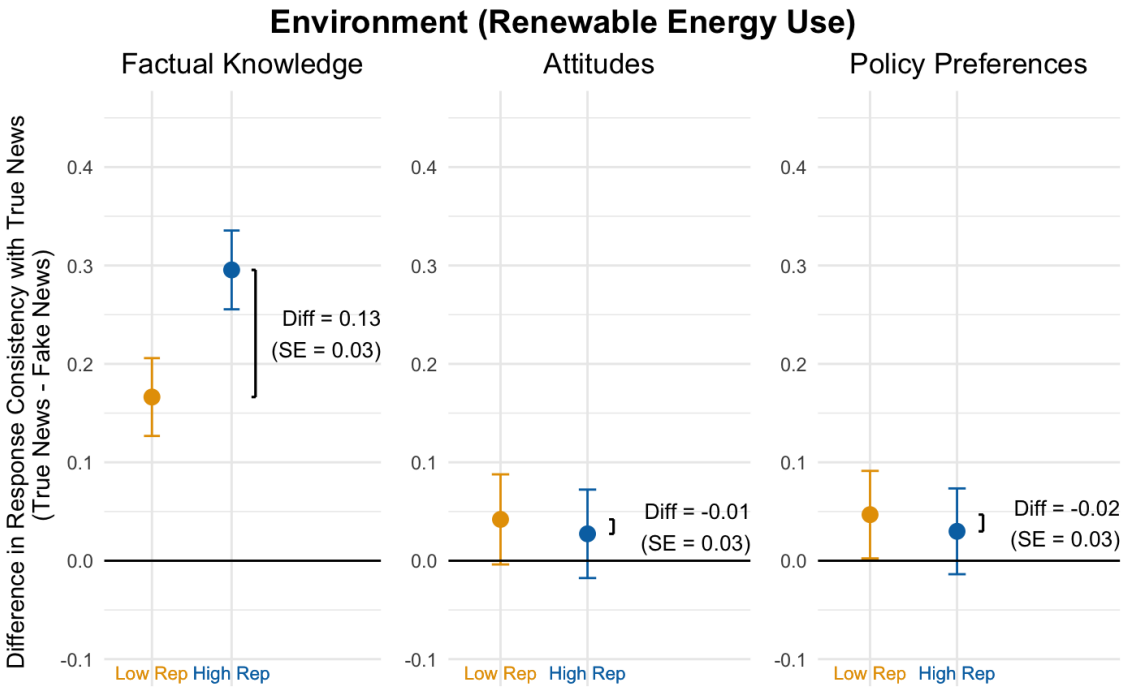

Figure 4: Effects of News about Energy and the Environment on Knowledge, Attitudes, and Policy preferences

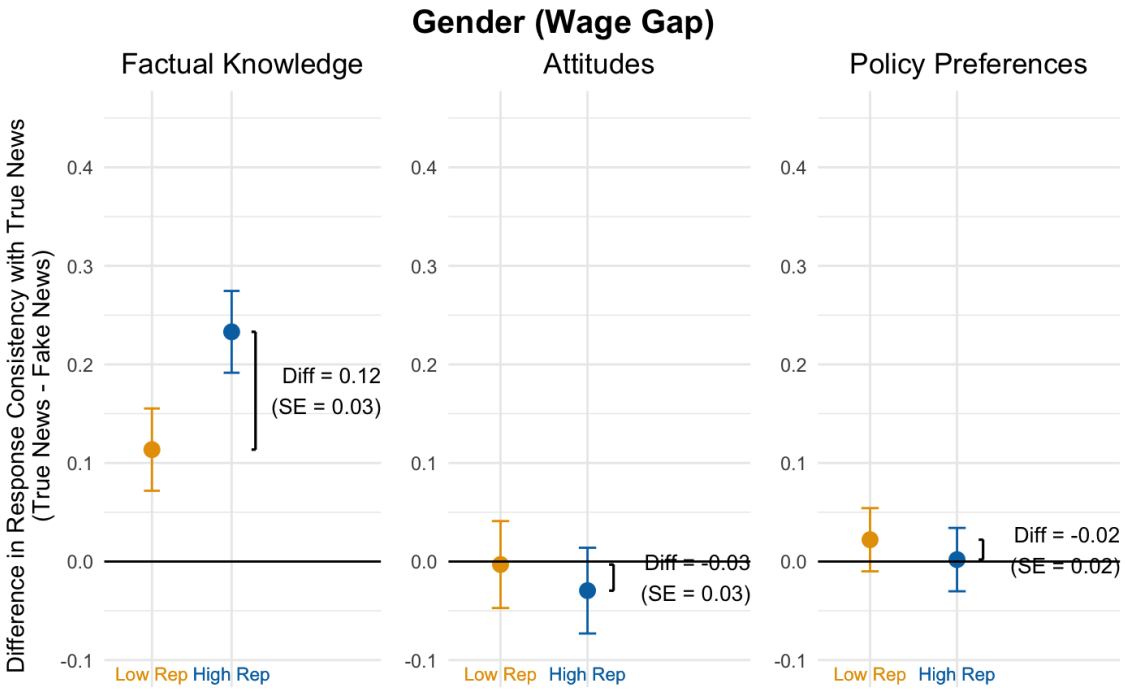

Figure 5: Effects of News about Gender Differences in Income on Knowledge, Attitudes, and Policy preference

s

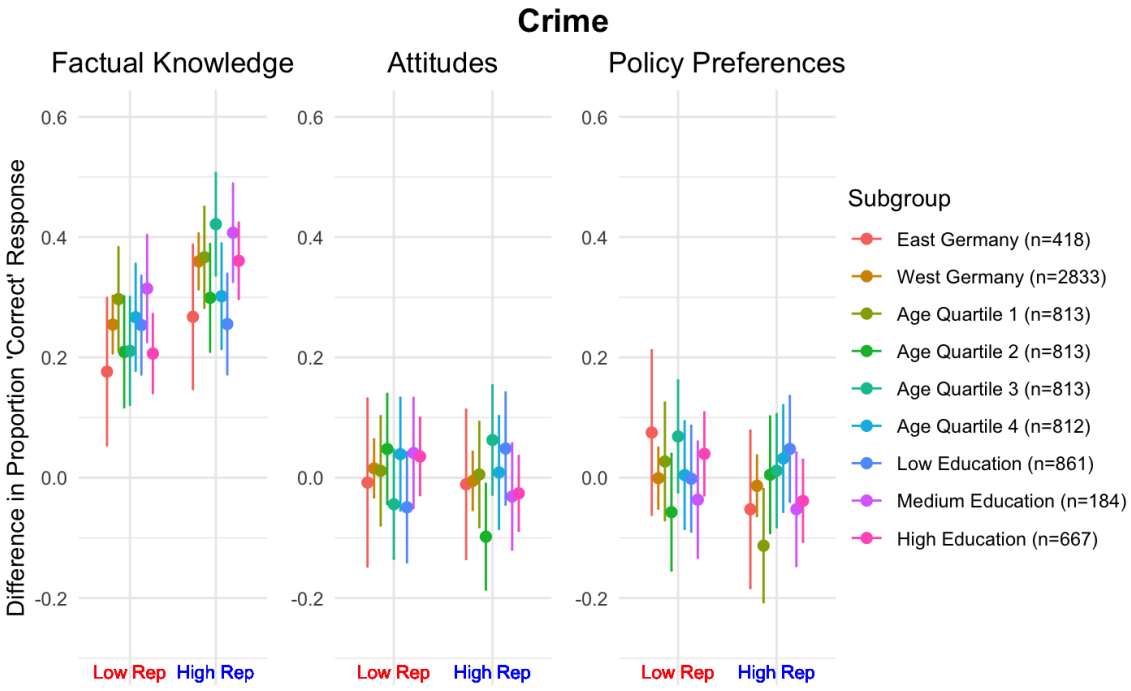

Figure 6: Subgroup analyses, breaking results in Figure 2 down by background characteristics. See legend for subgroups.

for false positive or multiple testing issues). We conclude that the average treatment effects we estimated may well be closer to constant treatment effects across all individuals, at least with respect to the background variables we measured.

The Supplementary Appendix includes analogous figures for unemployment, energy, and gender income equity, all giving the same conclusions about minimal differences across subgroups.

6 Implications

The stated purpose of most traditional, high reputability media outlets and the elaborate editorial operations they run is conveying factual knowledge about the world and, as communication portals, causing readers to understand the reported facts in the same way. Traditional outlets regularly convey that they have no intent to influence readers' political attitudes or policy preferences, except inasmuch as these views may be affected by the facts. Although editorial policies vary (*Focus Online*, in particular, is an avowedly nonpartisan, nonideological outlet with a reputation for being an approximately center or

center-right publication), opinion pages are under editorial control but clearly segregated from the news. Thus, our initial empirical results, the first that test these hypotheses directly, might be just what high reputation outlets have intended all along: The outlets are having remarkably large effects on readers' understandings of the facts and no noticeable effects on their political attitudes or policy preferences. So far so good.

Unfortunately, even if the editorial function of traditional media outlets are operating as they intended, they have now been joined in the broader news ecosystem by numerous upstart media outlets — groups or individuals that create low cost websites or publications on their own, all without any editorial infrastructure, assurance that their news stories are accurate, or claim to reputability. And, probably because so much of their revenue has been taken by modern technology companies, these major outlets now routinely sell access to their newspages and their readers to the highest bidder, including to low reputation content creators. These sales would cause no problem if sponsored content without editorial control had no impact on readers, but we show this is incorrect: They have a large effect, and only a modestly smaller effect on readers' factual knowledge than the high reputability editorial process itself, and they have the same (zero) effect on readers' political attitudes and policy preferences.

Our results confirm that traditional media outlets are indeed the crucial positive contributors to the news ecosystem, and likely to democracy, that responsible journalists claim. However, the remarkable impact of content sponsors without expensive editorial operations that we find is only slightly smaller than the traditional outlets. Of course, most sponsored content still seems to have commercial rather than policy motives and is only impactful by building off the reputable high quality outlet's brand.

7 Suggestions for Future Research

We encourage future researchers to pursue similar relationships with professionals in the news media industry so both sectors can together learn more about how the news media may impact the operation of democracies. Fully randomized experiments will help us avoid the severe endogeneity problems endemic in many observational studies of media

effects. Better relationships between academia and the news media industry should also enable scholars to estimate causal effects, and build understanding across, a much wider array of situations and sites than the one large site to which we obtained access for this experiment. How high and low reputability sites interact is also crucial, through advertisements as in our experiment or when they convince more well known sites, such as in social or traditional media, to amplify their message. Finally, the unique partnership that enabled this field experiment also required design constraints that we hope future work can relax or extend. Promising directions include assessing longer-term effects to determine whether impacts persist, fade, or amplify over time, and examining how outcomes vary across different institutional contexts.

References

Allen, Jennifer, Duncan J. Watts, and David G. Rand (2024). "Quantifying the impact of misinformation and vaccine-skeptical content on Facebook". In: *Science* 384.6699, eadk3451. DOI: 10.1126/science.adk3451.

Altay, Sacha, Manon Berriche, and Alberto Acerbi (2023). "Misinformation on misinformation: Conceptual and methodological challenges". In: *Social media+ society* 9.1, p. 20563051221150412.

Ansolabehere, Stephen D, Shanto Iyengar, and Adam Simon (1999). "Replicating experiments using aggregate and survey data: The case of negative advertising and turnout". In: *American political science Review* 93.4, pp. 901–909.

Bachmann, Philipp, Mark Eisenegger, and Diana Ingenhoff (2022). "Defining and measuring news media quality: Comparing the content perspective and the audience perspective". In: *The International Journal of Press/Politics* 27.1, pp. 9–37.

Balint, Anat (2023). ""It's in the Air"—Sponsored Editorial Content as a Path for Stealth Government Propaganda: The Case of Israeli Media". In: *Sponsored Editorial Content in Digital Journalism*. Routledge, pp. 88–109.

Banerjee, Abhijit, Nils Enevoldsen, Rohini Pande, and Michael Walton (2024). "Public information is an incentive for politicians: Experimental evidence from Delhi elections". In: *American Economic Journal: Applied Economics* 16.3, pp. 323–353.

Barabas, Jason and Jennifer Jerit (2010). "Are survey experiments externally valid?" In: *American Political Science Review* 104.2, pp. 226–242.

Barnes, Lucy and Timothy Hicks (2018). "Making austerity popular: The media and mass attitudes toward fiscal policy". In: *American Journal of Political Science* 62.2, pp. 340–354.

Barrera, Oscar, Sergei Guriev, Emeric Henry, and Ekaterina Zhuravskaya (2020). "Facts, alternative facts, and fact checking in times of post-truth politics". In: *Journal of public economics* 182, p. 104123.

Bartels, Larry M (1993). "Messages received: the political impact of media exposure." In: *American Political Science Review* 87.02, pp. 267–285.

Bauer, Paul C and Bernhard Clemm von Hohenberg (2021). "Believing and sharing information by fake sources: An experiment". In: *Political Communication* 38.6, pp. 647–671.

Berinsky, Adam J (2023). "Political Rumors: Why We Accept Misinformation and How to Fight It". In.

Bouton, Laurent, Julia Cagé, Edgard Dewitte, and Vincent Pons (2022). *Small campaign donors*. Tech. rep. National Bureau of Economic Research.

Brader, Ted, Nicholas A Valentino, and Elizabeth Suhay (2008). "What triggers public opposition to immigration? Anxiety, group cues, and immigration threat". In: *American Journal of Political Science* 52.4, pp. 959–978.

Buvár, Ágnes, Eszter Balogh, Attila Marcelián Dobai, and Judit Tessényi (2024). "Do You Care to Share? The Negative Effect of Sponsored Content on the Behavioral Engagement with a Social Media Post Containing a COVID-19 Message". In: *Journal of Interactive Advertising* 24.2, pp. 143–155.

Cagé, Julia and Valeria Rueda (2016). "The long-term effects of the printing press in sub-Saharan Africa". In: *American Economic Journal: Applied Economics* 8.3, pp. 69–99.

Campbell, Margaret C, Gina S Mohr, and Peeter WJ Verlegh (2013). "Can disclosures lead consumers to resist covert persuasion? The important roles of disclosure timing and type of response". In: *Journal of Consumer Psychology* 23.4, pp. 483–495.

Carlson, Matt (2015). "When news sites go native: Redefining the advertising–editorial divide in response to native advertising". In: *Journalism* 16.7, pp. 849–865.

Colliander, Jonas and Susanna Erlandsson (2015). "The blog and the bountiful: Exploring the effects of disguised product placement on blogs that are revealed by a third party". In: *Journal of Marketing Communications* 21.2, pp. 110–124.

Cooper, Christopher A and Anthony J Nownes (2004). "Money well spent? An experimental investigation of the effects of advertorials on citizen opinion". In: *American Politics Research* 32.5, pp. 546–569.

Costera Meijer, Irene and Hildebrand P Bijleveld (2016). "Valuable journalism: Measuring news quality from a user's perspective". In: *Journalism studies* 17.7, pp. 827–839.

Culpepper, Pepper D, Jae-Hee Jung, and Taeku Lee (2024). "Banklash: How media coverage of bank scandals moves mass preferences on financial regulation". In: *American Journal of Political Science* 68.2, pp. 427–444.

DellaVigna, Stefano and Ethan Kaplan (2007). "The Fox News effect: Media bias and voting". In: *The Quarterly Journal of Economics* 122.3, pp. 1187–1234.

Djourelova, Milena (2023). "Persuasion through slanted language: Evidence from the media coverage of immigration". In: *American economic review* 113.3, pp. 800–835.

Gaozhao, Dongfang (2021). "Flagging fake news on social media: An experimental study of media consumers' identification of fake news". In: *Government Information Quarterly* 38.3, p. 101591.

Gentzkow, Matthew, Edward L Glaeser, and Claudia Goldin (2006). "The rise of the fourth estate. How newspapers became informative and why it mattered". In: *Corruption and reform: Lessons from America's economic history*. University of Chicago Press, pp. 187–230.

Gentzkow, Matthew, Jesse M Shapiro, and Michael Sinkinson (2011). "The effect of newspaper entry and exit on electoral politics". In: *The American Economic Review* 101.7, pp. 2980–3018.

Gerber, Alan S, Dean Karlan, and Daniel Bergan (2009). "Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions". In: *American Economic Journal: Applied Economics* 1.2, pp. 35–52.

Gilbert, Daniel T, Elizabeth C Pinel, Timothy D Wilson, Stephen J Blumberg, and Thalia P Wheatley (1998). "Immune neglect: a source of durability bias in affective forecasting." In: *Journal of personality and social psychology* 75.3, p. 617.

Guriev, Sergei, Emeric Henry, Théo Marquis, and Ekaterina Zhuravskaya (2023). "Curtailing false news, amplifying truth". In: DOI: https://dx.doi.org/10.2139/ssrn.4616553.

Hayes, Danny and Jennifer L Lawless (2015). "As local news goes, so goes citizen engagement: Media, knowledge, and participation in US House Elections". In: *The Journal of Politics* 77.2, pp. 447–462.

Henry, Emeric, Ekaterina Zhuravskaya, and Sergei Guriev (2022). "Checking and sharing alt-facts". In: *American Economic Journal: Economic Policy* 14.3, pp. 55–86.

Hopkins, Daniel J, John Sides, and Jack Citrin (2019). "The muted consequences of correct information about immigration". In: *The Journal of Politics* 81.1, pp. 315–320.

King, Gary, Benjamin Schneer, and Ariel White (2017). "How the news media activate public expression and influence national agendas". In: *Science* 358.6364, pp. 776–780.

Klein, Tim, Elisabeth Fondren, and Leonard M Apcar (2019). "News editing and the editorial process". In: *Oxford Research Encyclopedia of Communication*.

Ladd, Jonathan McDonald and Gabriel S Lenz (2009). "Exploiting a rare communication shift to document the persuasive power of the news media". In: *American Journal of Political Science* 53.2, pp. 394–410.

Larreguy, Horacio, John Marshall, and James M Snyder Jr (2020). "Publicising malfeasance: when the local media structure facilitates electoral accountability in Mexico". In: *The Economic Journal* 130.631, pp. 2291–2327.

Lazer, David MJ, Matthew A Baum, Yochai Benkler, Adam J Berinsky, Kelly M Greenhill, Filippo Menczer, Miriam J Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. (2018). "The science of fake news". In: *Science* 359.6380, pp. 1094–1096.

Mercier, Hugo (2020). *Not born yesterday: The science of who we trust and what we believe*. Princeton University Press.

— (2022). "Confirmation bias–myside bias". In: *Cognitive illusions*. Routledge, pp. 78–91.

Muhlmann, Géraldine (2010). *Journalism for democracy*. Polity.

Pickard, Victor (2019). *Democracy without journalism?: Confronting the misinformation society*. Oxford University Press.

Porlezza, Colin (2019). "Accuracy in journalism". In.

Puglisi, Riccardo and James M Snyder (2015). "Empirical studies of media bias". In: *Handbook of media economics*. Vol. 1. Elsevier, pp. 647–667.

Reeves, Aaron, Martin McKee, and David Stuckler (2016). "'It's The Sun Wot Won It': Evidence of media influence on political attitudes and voting from a UK quasi-natural experiment". In: *Social science research* 56, pp. 44–57.

Schudson, Michael (2001). "The objectivity norm in American journalism". In: *Journalism* 2.2, pp. 149–170.

Snyder Jr, James M and David Strömberg (2010). "Press coverage and political accountability". In: *Journal of political Economy* 118.2, pp. 355–408.

Sperber, Dan, Fabrice Clement, Christophe Heintz, Olivier Mascaro, Hugo Mercier, Gloria Origgi, and Deirdre Wilson (2010). "Epistemic vigilance". In: *Mind & language* 25.4, pp. 359–393.

Stiglitz, Joseph E, Anya Schiffrin, and Dylan W Groves (2024). "Journalism for development: the role of journalism promoting democracy and political accountability and sustainable development". In.

Thaler, Michael (2024). "The fake news effect: Experimentally identifying motivated reasoning using trust in news". In: *American Economic Journal: Microeconomics* 16.2, pp. 1–38.

Van Reijmersdal, Eva A, Marieke L Fransen, Guda Van Noort, Suzanna J Opree, Lisa Vandeberg, Sanne Reusch, Floor Van Lieshout, and Sophie C Boerman (2016). "Effects of disclosing sponsored content in blogs: How the use of resistance strategies mediates effects on persuasion". In: *American Behavioral Scientist* 60.12, pp. 1458–1474.

Vavreck, Lynn (2007). "The exaggerated effects of advertising on turnout: The dangers of self-reports". In: *Quarterly Journal of Political Science* 2.4, pp. 325–343.

Vavreck, Lynn and Shanto Iyengar (2011). "The future of political communication research". In: *The Oxford handbook of American public opinion and the media*. Oxford University Press Oxford, pp. 156–168.

Vehkoo, Johanna (2010). "What is quality journalism and how it can be saved". In: *Reuters Institute for the Study of Journalism*, pp. 1–76.

Watts, Duncan J and David M Rothschild (2017). "Don't blame the election on fake news. Blame it on the media". In: *Columbia Journalism Review* 5, pp. 67–84.

Wojdynski, Bartosz W (2016). "Native advertising: Engagement, deception, and implications for theory". In: *The new advertising: Branding, content and consumer relationships in a data-driven social media era*, pp. 203–236.

Wojdynski, Bartosz W and Nathaniel J Evans (2016). "Going native: Effects of disclosure position and language on the recognition and evaluation of online native advertising". In: *Journal of Advertising* 45.2, pp. 157–168.

Wood, Thomas and Ethan Porter (2019). "The elusive backfire effect: Mass attitudes' steadfast factual adherence". In: *Political Behavior* 41, pp. 135–163.

Woolley, Samuel C and Philip N Howard (2018). *Computational propaganda: Political parties, politicians, and political manipulation on social media*. Oxford University Press.

Zeng, Eric, Tadayoshi Kohno, and Franziska Roesner (2020). "Bad news: Clickbait and deceptive ads on news and misinformation websites". In: *Workshop on Technology and Consumer Protection*, pp. 1–11.