Untrustworthy Website Exposure and Election Beliefs: Selective Exposure and Ideological Asymmetry

Ross Dahlke and Jeffrey T. Hancock

Abstract. The proliferation of untrustworthy information, particularly during election periods, poses a significant challenge to democratic processes worldwide. Combining a two-wave panel study with web-browsing behavior (N = 21M) of 1,194 American adults around the 2020 US presidential election, we examine associations between visits to untrustworthy websites at the domain level and subsequent election beliefs. We find that those exposed to untrustworthy websites are 17.3% more likely to believe that the election was not won by the certified winner in a post-election survey. However, one challenge in assessing the relationship between untrustworthy website exposure and beliefs is disentangling the behavior of individuals selectively exposing themselves to information congenial to an existing belief versus prior partisanship heterogeneously relating to differing changes in beliefs. To address this challenge, we apply a machine learning method that controls for the propensity to be exposed to untrustworthy websites. Using this method, we find that exposure to untrustworthy websites is associated with the likelihood of believing that the certified election winner did not win by 4.2%. This association, however, is asymmetric: For conservatives, whose candidate lost the election, the association is 12.6%, while for liberals, the association is negligible at -0.2%. Further, we identify a dosage relationship—each additional exposure to untrustworthy websites increases the association between untrustworthy website exposure and election outcome beliefs among conservatives by .035% but only .010% for liberals. Through a series of stringent robustness checks, we find that these directional associations are unlikely to be "explained away" by potential unobserved variables, adding credibility to the estimates. Our results underscore the importance of addressing untrustworthy information not as a uniform influence but as a factor whose relationship with beliefs is contingent upon environmental context, individual predispositions, and exposure volume.

doi:10.54501/jots.v3i1.250

1 Introduction

Among global leaders, misinformation (information that is false, misleading, or unsubstantiated (Nyhan and Reifler 2010)) was recently listed as society's biggest problem (World Economic Forum 2024). Academic research provides a more nuanced understanding of this issue (Freelon and Wells 2020; Weeks and Gil de Zúñiga 2021), finding that exposure is generally "Less Than You Think" (Guess, Nagler, and Tucker 2019), constitutes a relatively small portion of people's information diets (Allcott and Gentzkow 2017; Allen et al. 2020; Altay, Nielsen, and Fletcher 2022), and is declining over time (Moore, Dahlke, and Hancock 2023). Nonetheless, tens of millions of American adults are exposed to such websites (Guess, Nyhan, and Reifler 2020; Moore, Dahlke, and Hancock 2023; Zhou, Yang, and González-Bailón 2025). Exposure is also highly concentrated, with a small number of people accounting for the majority of exposures or clicks (Grinberg et al. 2019; Guess, Nagler, and Tucker 2019; Nelson and Taneja 2018), suggesting that although exposure to such information may not be as widespread as public discourse implies, it may still play a role in democratic processes for at least some subset of the population (Robertson 2022; Ecker et al. 2024).

While these descriptive findings are important to understanding the reach and distribution of untrustworthy information, they typically do not shed light on the real-world relationship between exposure to such information and beliefs. That is, what are the societal consequences of exposure to untrustworthy websites? Answering this question is crucial, as many contemporary fears about democratic health are rooted in an assumption that this exposure has a powerful, direct effect on its audience (Budak et al. 2024; Altay, Hoes, and Wojcieszak 2025).

To address this concern, we combine a two-wave panel study with web-browsing behavior data (N = 21M observations) from 1,194 American adults around the 2020 US presidential election. We apply a nonparametric double machine learning method that explicitly models and controls for selective exposure—the tendency for individuals to seek information congruent with existing beliefs—while estimating the association between visits to untrustworthy websites at the domain level and subsequent election beliefs. This approach allows us to disentangle the propensity to be exposed to untrustworthy websites from the relationship between exposure and belief formation, and enables examination of heterogeneous associations across partisan groups. We measure election beliefs through a direct question about who participants believe won the 2020 election, asked approximately one month after the election during a critical period of emerging fraud narratives (Dahlke and Pan 2024; Starbird, DiResta, and DeButts 2023).

Our findings reveal three key results. First, after controlling for selective exposure, those exposed to untrustworthy websites are 4.2% more likely to believe that the certified election winner did not actually win—a substantial reduction from the 17.3% raw difference, indicating that selective exposure accounts for approximately 75% of the observed correlation. Second, this association is markedly asymmetric (i.e., moderated by participant ideology): Conservatives show a 12.6% association while liberals show essentially no association (-0.2%), suggesting that information effects are contingent upon the alignment between content and recipient predispositions rather than operating uniformly. Third, we identify a dosage relationship where each additional exposure to untrustworthy websites increases the association by 0.034% overall, with this effect again showing asymmetry (conservatives: 0.035%; liberals: 0.01%). These results show that the relationship between untrustworthy information exposure and democratic beliefs is not a simple "hypodermic needle" effect (Bineham 1988; Bello-Pardo 2022), but rather depends critically on individual characteristics and environmental context.

2 Literature Review

2.1 Experimental Evidence on Information Exposure

Experimental research in the political domain shows that untrustworthy information can meaningfully alter beliefs and preferences, though the strength and scope of these effects vary across contexts. Individuals have been shown to form false memories about political events when exposed to fabricated news stories (Murphy et al. 2019) or doctored images (Frenda et al. 2013). Exposure to false information about election fraud can increase support for postponing elections (Craig and Gainous 2024), and even a single exposure to false political information may boost voting intention, though such exposures show limited associations with other political beliefs and behaviors (Guess, Nyhan, and Reifler 2020). Related work in other domains has tended to find more modest effects: Exposure to vaccine misinformation shows small associations with intentions and behaviors (Allen, Watts, and Rand 2024; Featherstone and Zhang 2020; Greene et al. 2022; Saint Laurent et al. 2022), and exposure to climate change skepticism has shown limited impact on attitudes (Drummond, Siegrist, and Árvai 2020; Zhou and Shen 2022).

A growing body of field experiments manipulating social media users' informational environments provides robust evidence for the limited impact of platform-level changes on political beliefs. Studies altering content availability—by removing political ads or all reshared content, or by reducing content from like-minded sources—find that while exposure patterns change, core political attitudes and beliefs remain stable (Allcott et al. 2025; Guess et al. 2023b; Nyhan et al. 2023). Even more fundamental changes to the information architecture, such as switching from an algorithmic to a chronological feed, limiting access to multimedia content in social media, or deactivating social media access entirely, similarly fail to produce significant shifts in political knowledge, polarization, or belief in misinformation (Allcott et al. 2024; Arceneaux, Zengin, and Ladd, n.d.; Guess et al. 2023a; Ventura et al. 2023). Likewise, interventions designed to contextualize information, such as adding credibility labels or incentivizing exposure to partisan news, also show minimal effects on political opinions or affect, though they can produce secondary effects like reduced trust in mainstream media (Aslett et al. 2022; Guess et al. 2021).

Taken together, this diverse set of environmental manipulations highlights the resilience of political beliefs against top-down changes in the information ecosystem, though some scholars argue that the limited effects observed in individual-level experiments may reflect methodological constraints rather than the absence of meaningful platform influence on complex social phenomena (Bak-Coleman et al. 2025). This resilience, however, does not imply that the content within these environments is without consequence. Indeed, the composition of the information users see can produce important, if subtle, effects. For instance, consuming high-quality news can enhance knowledge and improve truth discernment, while exposure to higher proportions of false news can erode media trust and inflate overconfidence in one's discernment abilities, even while leaving core attitudes unchanged (Altay, Hoes, and Wojcieszak 2025; Altay, Lyons, and Modirrousta-Galian 2024). This suggests that while broad architectural interventions may be blunt instruments, the specific nature of information exposure remains a critical factor in shaping more nuanced civic outcomes.

2.2 Observational Evidence and Methodological Challenges

Observational studies find significant, albeit limited, correlations between untrustworthy information exposure and the outcome of interest. For example, past work has found a positive correlation between exposure to misinformation websites and false beliefs, but not intentions to participate in politics (Guess, Nyhan, and Reifler 2020). There is also a correlation between engaging with election-related misinformation on Twitter and voter turnout (Green et al. 2022). Those who shared misinformation on Twitter promoting election fraud misinformation were less likely to vote. However, other work has not found a strong relationship between visiting untrustworthy websites and belief in political misperceptions (Weeks and Gil de Zúñiga 2021). These documented correlations, while informative, raise concerns about issues like reverse causality (Adams et al. 2023; Guess and Lyons 2020; Petersen 2023).

2.3 Theoretical Frameworks: Selective Exposure and Asymmetry

Conceptually, simple correlations between exposure and outcome conflate selective exposure (Sears and Freedman 1967; Stroud 2008, 2010; Zillmann and Bryant 2013) and asymmetric differences in antecedents to exposure. Selective exposure refers to a tendency for people who already hold certain beliefs to seek out consistent information (Sears and Freedman 1967; Zillmann and Bryant 2013). For example, conservative individuals are more likely to expose themselves to conservative-leaning untrustworthy websites and liberal individuals to liberal-leaning untrustworthy websites (Guess, Nyhan, and Reifler 2020; Moore, Dahlke, and Hancock 2023). One possible explanation for asymmetric partisan differences is motivated reasoning, which refers to how people are motivated to arrive at or justify specific conclusions to new information based on desired beliefs (Epley and Gilovich 2016; Kunda 1990), ultimately accepting information consistent with desired beliefs and rejecting contradictory information (Druckman and McGrath 2019). Alternative explanations for these asymmetric effects exist. For instance, work on ideological differences (Jost et al. 2018) suggests fundamental psychological differences between conservatives and liberals in information processing and belief formation. While most observational data cannot definitively distinguish between these mechanisms, both theoretical frameworks predict asymmetric effects.

The distinction between selective exposure and post-exposure effects allows us to sharpen this theoretical question. By controlling for the first stage (selective exposure), we can test competing hypotheses about the second. One hypothesis, echoing "hypodermic needle" theories, is that the effect of exposure is *symmetrical*. This account predicts uniformity across partisan groups once they receive the information. An alternative hypothesis, grounded in the theories of motivated reasoning and ideological differences mentioned above, is that the effect is *asymmetrical*. This would mean that even after accounting for selection, the effect of the information remains conditional on partisanship, primarily influencing the group for whom the message is psychologically congenial. Our study is designed to adjudicate between these two possibilities.

Taken together, the existing literature presents a mixed picture. While experiments demonstrate that exposure to untrustworthy information can influence beliefs under controlled conditions, observational studies reveal the complexities of real-world exposure patterns and raise persistent challenges for causal inference. Traditional regression approaches, commonly used in these observational studies, struggle to account for the fact that both information exposure and its effects on beliefs are likely to be heterogeneous and shaped by prior attitudes (Slater 2007; Stroud 2010). Theoretical frameworks like selective exposure and motivated reasoning predict asymmetric responses to information, yet rigorously testing these asymmetries remains difficult without flexible methods that can model complex selection processes and interactions.

2.4 Present Study

This study aims to bridge these gaps by leveraging large-scale web-browsing data combined with panel surveys. We employ double machine learning (DML) to explicitly model and control for selective exposure to untrustworthy website domains, thereby allowing for a clearer estimation of the association between visits to untrustworthy websites and election-related beliefs. This approach also enables us to examine heterogeneity in effects across partisan groups, bringing methodological tools into better alignment with the theoretical expectations that guide contemporary research on misinformation.

To address these challenges, this study leverages large-scale web-browsing data combined with panel surveys and applies a nonparametric double machine learning approach. This design enables more rigorous estimation of the association between visits to untrustworthy websites and election-related beliefs, while directly testing for theoretically predicted heterogeneity in effects across partisan groups. Double machine learning is particularly well-suited for this task, as it addresses key theoretical and methodological concerns in three ways. First, its nonparametric modeling of exposure propensity captures the complex, nonlinear interactions among individual characteristics that shape selective exposure (Bennett and Iyengar 2008). Second, its flexible estimation of treatment heterogeneity reflects theoretical expectations that information effects vary across subpopulations based on factors such as prior beliefs and identity-protective cognition (Kahan 2017). Third, by discovering rather than prespecifying interaction patterns, DML allows the data to reveal how belief-updating processes may differ across individuals, rather than imposing rigid assumptions in advance.

Our measurement strategy prioritized capturing time-sensitive beliefs about election legitimacy during a crucial period of democratic discourse. To measure beliefs about the election outcome, about a month after the 2020 US presidential election, we asked participants who they believe won the election. The survey was conducted in December 2020—a unique window between Election Day in November and the January 6 events—capturing beliefs during the emergence of election fraud narratives, before they became entrenched in broader political discourse. Our straightforward question directly assesses the core belief at issue, and its timing provides unique insight into the early formation of these beliefs. The measure's simplicity and strategic timing enabled us to document this rapidly evolving phenomenon during a pivotal moment in its development.

In this study, we investigate three key questions about the relationship between untrustworthy website exposure and election beliefs: (1) What is the overall association between exposure to untrustworthy websites and beliefs about the 2020 US presidential election outcome after controlling for selective exposure? (2) Does this association differ between supporters of the winning versus losing candidate? (3) Is there a dosage effect where additional exposures strengthen this relationship? Our findings reveal that exposure to untrustworthy websites is associated with a 4.2% greater likelihood of believing the certified winner did not win, with substantial asymmetry between conservatives (12.6%) and liberals (-0.2%), and a cumulative dosage effect that varies by political predisposition. These results suggest that the relationship between untrustworthy information and beliefs is contingent upon both individual characteristics and environmental context.

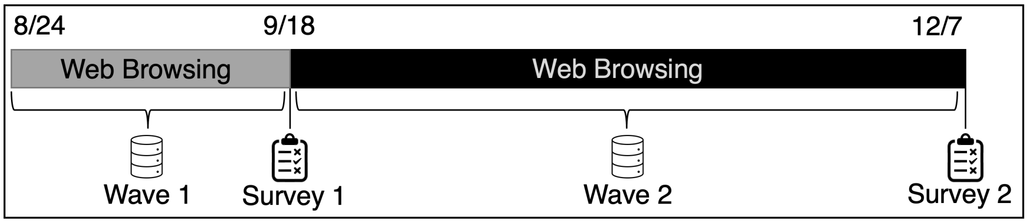

Figure 1: Timeline of data collection. The first wave of web-browsing data was collected from August 24, 2020, to September 18, 2020. We administered a survey on September 18, 2020, to collect demographic information and political support. The second wave of web-browsing data was collected from September 19, 2020, to December 7, 2020. Then, we conducted a second survey on December 7, 2020, and asked participants about their beliefs on whether the election was not won by the certified winner.

3 Untrustworthy Website Exposure and False Belief Measurement

We collected data from 1,194 participants recruited by the survey firm YouGov. These participants completed Survey 1 on September 18, 2020, seven weeks before the US presidential election, and Survey 2 on December 7, 2020, four weeks after the presidential election. We also gathered two waves of web-browsing data from the participants using YouGov's Pulse web-tracking software. The first wave of browsing data was collected from August 24, 2020, to September 18, 2020, and the second wave of web-browsing data was collected from September 19, 2020, to December 7, 2020. Figure 1 summarizes the data collection process.

All participants consented to the surveys and installed the web-tracking software, and YouGov compensated them for their participation. YouGov provided survey weights to match the sample to a nationally representative population, which we apply in our regression analyses to enhance generalizability (see Appendix A). The data and code supporting this study's findings are available in OSF with the identifier https://osf.io/egnwd/ (Dahlke and Hancock 2022). We complied with all relevant ethical regulations. Stanford University's Institutional Review Board (IRB) approved the study protocol.

To measure our main dependent variable, we asked participants, "In your view, who won the presidential election held in November?" on December 7, 2020, in Survey 2. Notably, December 7, 2020, was a month after the election and one month before the January 6 events, making this one of the first academic studies to measure this belief. In our survey, 19.7% (95% CI [17.4, 22.0]) of participants said they believed Donald Trump was the rightful winner, including 47.0% (95% CI [42.3, 51.8]) of Trump supporters. Importantly, other research has found that this belief is genuinely held and is not "expressive responding" to demonstrate partisan membership (Fahey 2023; Graham and Yair 2023).

The main independent variables of interest are untrustworthy website exposure and the number of untrustworthy website exposures in Wave 2 of the web-browsing data, both of which operate at the domain level. To identify participant visits to untrustworthy websites, we followed the procedure used in recent studies (Moore, Dahlke, and Hancock 2023; W. Ahmad et al. 2024) by creating a comprehensive list of 1,796 domains for the 2020 US election context. This procedure involves combining two main sources: the list of untrustworthy websites from Guess, Nyhan, and Reifler (2020) and an expanded list of domains from NewsGuard.[1] The base list from Guess, Nyhan, and Reifler (2020) covers the 2016 election and is a combination of 490 domains from Grinberg et al. (2019) and 66 from Allcott and Gentzkow (2017). To this, we added 1,240 domains that NewsGuard, an organization of journalists who rate news sources, classified as "repeatedly publishing false content" during the 2020 study period. Our primary independent variable, binary untrustworthy website exposure, was coded "1" if a participant visited at least one unique domain from this list during the Wave 2 observation period (September 19, 2020, to December 7, 2020), and "0" otherwise. Any recorded page view to a URL within a listed domain constituted a visit. Our secondary independent variable, exposure dosage, was operationalized as the total count of page views across all listed domains during Wave 2 for each participant. We use a variety of variables as the observable control variables in our analyses, including partisanship (Trump supporter), education level, gender, race, political knowledge, political interest, and age. We also control for digital behaviors of exposure to untrustworthy websites in Wave 1 and the total number of websites visited.

4 Untrustworthy Website Exposure and Belief Relationship Estimation Method

We used a nonparametric double machine learning method developed by Wager and Athey (2018) to estimate the average association between untrustworthy website exposure and belief in the certified loser winning the election. This method flexibly adjusts for observed differences to estimate the average difference between people exposed to untrustworthy websites and those who were not. It is preferable to a simple OLS regression because it nonparametrically considers the propensity to be exposed based on the known variables of each individual, and these propensities are then incorporated when calculating the average relationship. Additionally, the method facilitates examination of heterogeneity in the results. For comparable results under a parametric regression framework, see Appendix A.

Like other double machine learning methods, the specific method we used (Wager and Athey 2018) is comprised of two main steps to estimate associations under a predicted outcomes framework. In the first machine learning step, we train a machine learning model to predict whether individual i is exposed to untrustworthy websites. For binary exposure, the methodology uses observed variables X for each individual and W, indicating whether that individual was exposed, to estimate the propensity of exposure,

1. https://www.newsguardtech.com/.

$\hat{W}$. In our case

$$X_i = \begin{bmatrix} Partisanship_i \\ PoliticalKnowledge_i \\ PoliticalInterest_i \\ Female_i \\ EduHSorLess_i \\ EduSomeCollege_i \\ EduCollegeGrad_i \\ EduPostGrad_i \\ RaceBlack_i \\ RaceHispanic_i \\ RaceWhite_i \\ RaceOther_i \\ AgeUnder30_i \\ Age30to44_i \\ Age45to65_i \\ Age65plus_i \\ ExposedUntrustworthyWave1_i \\ WebVisitsWave1_i \end{bmatrix}$$

and W_i = ExposedUntrustworthyWave2$_i$, with W_i being "0" if not exposed to at least one untrustworthy website and "1" if exposed. This first machine learning step is flexible regarding the method used to estimate $\hat{W}$. For example, one could use a simple logistic regression or a more complex neural network to predict exposure. In this manuscript, we follow Wager & Athey's (2018) methodology and use random forests with 2,000 trees and a target minimum node size of 5, and we use 50% of the data used to build each tree and "honest" fitting (see Wager and Athey (2018) for more details on these parameters).

In the second machine learning step, the observed variables X are considered along with $\hat{W}$. Our estimation strategy targets the overlap population—the subset of individuals for whom there is strong common support in the data (i.e., similar propensity scores for exposure regardless of actual exposure status)—to prioritize the internal validity of the estimate. The specific method we use, suggested by Li, Morgan, and Zaslavsky (2018), achieves this goal by focusing on observations where the estimated propensity of exposure is not close to 0 or 1, a practice that improves the credibility of the estimate by ensuring treatment and control units are comparable. Consistent with this focus on internal validity, we do not use the survey weights when estimating the causal forest, as the goal of the propensity model is to achieve balance within the sample, not to make generalizable inferences about the population (Zanutto 2006; DuGoff, Schuler, and Stuart 2014). The resulting association, $\tau(x)$, represents the average association for individuals with "moderate" propensity scores for exposure (i.e., those who have neither very high nor very low predicted probabilities of visiting untrustworthy websites). This focus on the overlap population enhances internal validity but means our estimates may not generalize to individuals with extreme exposure propensities. More formally, this all comes together as

$$\tau(x) = \frac{E\left[e(X)(1 - e(X))(Y^{(1)} - Y^{(0)})\right]}{E\left[e(X)(1 - e(X))\right]}, \tag{1}$$

where

$$e(x) = P(W_i = 1 \mid X_i = x). \tag{2}$$

Said in plain words, under the predicted outcomes framework we use the trained machine learning model to predict the outcome of interest (i.e., belief about who won the election) for each individual under two scenarios: a scenario in which the individual was exposed to untrustworthy websites in the second wave ($Y^{(1)}$) and a scenario in which they were not exposed ($Y^{(0)}$). The individual association is the difference between these two predictions. We then use each individual's likelihood to be exposed to untrustworthy websites (i.e., their propensity to selectively expose themselves to such websites, $\hat{W}$) as a weight to calculate the average association. To calculate the dosage relationship, we repeat this process but with W being the number of exposures to untrustworthy websites (i.e., not a binary variable).

5 Untrustworthy Website Exposure and Belief Relationship Estimation Results

While our main estimate of interest is the relationship between untrustworthy website exposure and beliefs about the results of the 2020 US presidential election, we describe the results of both steps of the machine learning pipeline. First, we examine Wave 2 untrustworthy website exposure. In our data, 39.6% (95% CI [36.8, 42.4]) of our participants were exposed to at least one untrustworthy website in Wave 2. While past studies have examined the predictors of exposure—for example, finding that older adults and conservatives are more likely to be exposed to untrustworthy information (Guess, Nyhan, and Reifler 2020; Moore, Dahlke, and Hancock 2023)—we approach this step as a prediction problem. The random forests used to predict exposure are well calibrated in a heteroskedasticity-consistent calibration test ($p < .001$; AUC = 0.754, see Appendix B), showing that the model accurately predicts Wave 2 exposure on held-out data. In examining variable importance (see Appendix C), Wave 1 exposure is the most "important" variable in predicting Wave 2 exposure. The next most important variables are political knowledge, partisanship, and political interest. This predicted propensity to be exposed to untrustworthy websites in Wave 2 is then considered as a weight, shown formally in Equation 1 and Equation 2 and visually in Appendix C. Conceptually, we argue that the incorporation of this first machine learning model into the second allows us to control for selective exposure (i.e., the predisposition to expose oneself to untrustworthy websites).

Next, we turn to the second machine learning model and use the resulting predictions with the predicted outcomes framework to examine the relationship between exposure and beliefs. After controlling for selective exposure, exposure to untrustworthy websites is associated with a 4.2% (95% CI [0.3, 8.0]) greater likelihood of believing the certified election winner did not win. Consistent with both motivated reasoning and ideological asymmetry accounts, we find a conditional average association of 12.6% (95% CI [1.7, 23.5]) for conservatives and -0.2% (95% CI [-1.9, 1.6]) for liberals (see Figure 2; for individual associations see Appendix D).

Finally, examining the relationship between exposure frequency and beliefs, we find an average estimated dosage association of 0.034% (95% CI [0.018, 0.049]), suggesting that each additional exposure to an untrustworthy website is associated with a 0.03% greater likelihood of believing the certified winner did not win. This relationship also shows asymmetry: The dosage association for conservatives was 0.035% (95% CI [0.017, 0.054]) and for liberals, 0.010% (95% CI [-0.020, 0.041]). These key findings—the positive

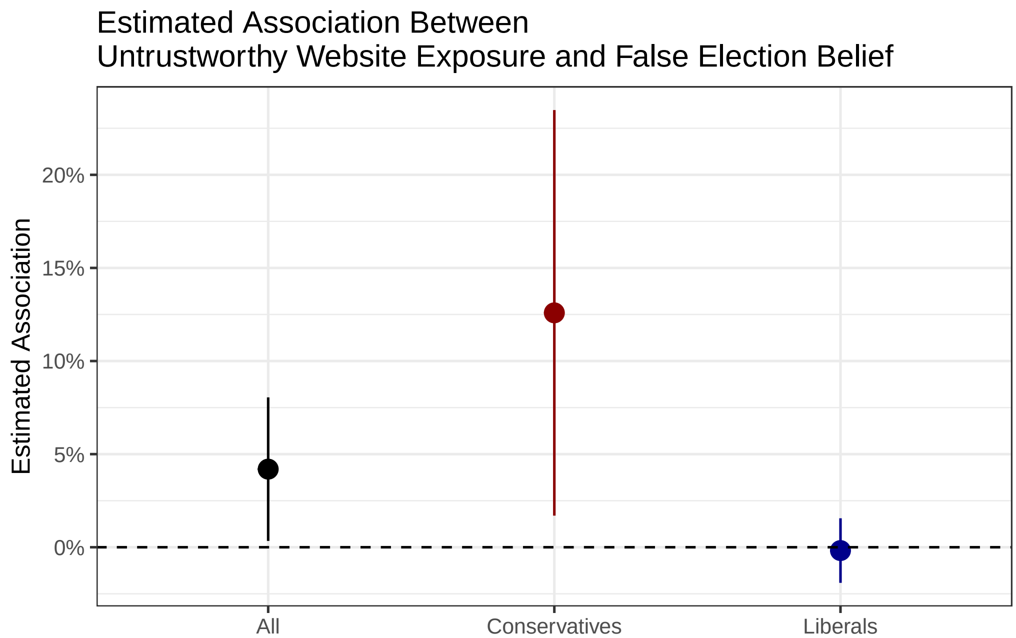

Figure 2: Plot of the average association and conditional average association and 95% confidence interval of untrustworthy website exposure on the belief that the losing candidate actually won. The x-axis represents different groups of people: all participants in our sample, conservatives, and liberals. The y-axis represents the average association of misinformation exposure on the belief that the certified election winner did not win for the given group.

overall association, the dosage effect, and the pronounced partisan asymmetry—are robust to alternative estimation strategies, including other double machine learning methods (Chernozhukov et al. 2018) and propensity score matching (see Appendix B for details).

6 Discussion

We examine the relationship between exposure to untrustworthy websites and beliefs about who won the 2020 US presidential election. After controlling for selective exposure via double machine learning, we find a significant association between exposure to untrustworthy websites and these beliefs. The 17.3% topline difference between those exposed to untrustworthy websites and those not decreases to a 4.2% difference when controlling for selective exposure. This result suggests that 75.7% of this topline difference can be explained by selective exposure, leaving 24.3% as the remaining difference.

Our findings adjudicate between the competing symmetrical and asymmetrical hypotheses. A symmetrical effect would have appeared as a positive association for both conservatives and liberals. Instead, we find a starkly asymmetrical effect: a large, positive association for conservatives (12.6%) and a null effect for liberals (-0.2%). This occurs even though liberals in our sample *were* exposed to these same domains, providing a real-world and continued refutation of "hypodermic needle" models of uniform media effects.

These associations align with prior work suggesting that exposure to untrustworthy information relates to various beliefs. The heterogeneity we find, that the losing party's supporters show stronger associations with believing the certified winner did not win, is consistent with both motivated reasoning research that finds partisans respond more positively to congenial information (Bello-Pardo 2022; Ditto, Liu, et al. 2019; Ditto, Clark, et al. 2019; Douglas et al. 2019; Kunda 1990; Nyhan 2020; Peterson and Iyengar 2021; Taber and Lodge 2006) and work suggesting differences in how partisans process information (Baron and Jost 2019; Federico, Deason, and Fisher 2012; Garrett and Bond 2021; Graham, Haidt, and Nosek 2009; Jost et al. 2018; Pennycook and Rand 2019; Van Bavel and Pereira 2018). This evidence contrasts concerns that untrustworthy information is a hypodermic needle that automatically changes beliefs upon exposure (A. K. Ahmad et al. 2022). Instead, we find that the differences in misinformation exposure associations are context-dependent. Instead, the associations we observe are likely felt in very specific contexts by only a select number of individuals about salient and congenial topics. For example, these findings align with work done on the 2016 and 2024 elections, which also document the development of conspiracy theories and conspiratorial beliefs around the validity of the elections (Center for an Informed Public, n.d.; Justwan, Baumgaertner, and Curtright 2024).

This heterogeneity is particularly important in the context of the dosage effect. Past work on misinformation effects has drawn on literature on the illusory truth effect (Pennycook, Cannon, and Rand 2018; Newman et al. 2020; Jalbert, Newman, and Schwarz 2020; Udry and Barber 2024; Vellani et al. 2023). This theory states that repeated exposure to information is more likely to cause a change than a single exposure (Ernst, Kühne, and Wirth 2017; Fazio et al. 2015; Hassan and Barber 2021). Indeed, the dosage is a key component in the story of misinformation effects. While some research suggests the illusory truth effect operates consistently across certain individual differences (Calvillo and Smelter 2020), other work documents meaningful individual variation in susceptibility to this effect, with some individuals even showing negative truth effects (Nadarevic and Erdfelder 2014). Additionally, contextual factors such as expectations about encountering falsehoods can moderate the effect (Jalbert, Newman, and Schwarz 2020). Our results contribute to this literature by demonstrating substantial heterogeneity in how repeated exposure to untrustworthy websites relates to belief formation, particularly when considering ideological congruence between the content and the recipient's predispositions. The heterogeneity in dosage effects we observe suggests that the mechanisms underlying repeated exposure effects may be more context-dependent and individual-specific than previously understood, operating most strongly when information aligns with motivated reasoning processes.

6.1 Limitations

Due to the observational nature of our data, the reported relationships should be interpreted as associations; inferring direct causality is subject to the limitations inherent in non-experimental research, particularly the untestable assumption of no unobserved confounding. There are a number of other limitations and assumptions that our results should be interpreted in light of. First, our measure of election beliefs uses a single item that asks who won the election. While this direct question captures the core belief of interest during a crucial historical moment, future work should employ validated multi-item scales that have since been developed. Second, because our outcome variable was measured only once post-election, we could not leverage panel data models like two-way fixed effects; future research that captures such beliefs at multiple time points could do so to further account for unobserved confounding. Third, our web-browsing data captures only one channel of exposure to untrustworthy information. People may encounter similar content through social media, messaging apps, or traditional media—channels we cannot measure in this study. Further, like other web-browsing studies, we analyze website visits at the domain level. This domain-level analysis means that not every URL from an untrustworthy website necessarily contains false or misleading information. From a theoretical standpoint, any visit to a flagged domain immerses users in a broader information ecosystem, exposing them to sidebars, recommendation widgets, and comment threads where election-fraud narratives may circulate (Anderson et al. 2014; Asker and Dinas 2017; Toepfl and Piwoni 2015). The entire domain, therefore, functions as a point of exposure to this ecosystem. Future work could build on our findings by employing more fine-grained, URL-level coding to distinguish exposure to verifiably false articles from other content. While there are difficulties in post hoc scraping of untrustworthy and hard news websites to capture the original content that users actually were exposed to (Dahlke et al. 2023), this direction represents a significant and emerging research frontier focused on understanding differential exposure to specific narratives (e.g., Goel et al. (2025) and Lavigne et al. (n.d.)). In addition, we examine whether the association between untrustworthy website exposure and election beliefs varies by the ideological slant of the domains themselves. Building on domain ratings from Robertson et al. (2018), we classify untrustworthy websites as either liberal or conservative and re-estimate our double machine learning models for these subsets (see Appendix E). The results indicate that the observed association is stronger for exposures to conservative-leaning untrustworthy websites, particularly among individuals whose predispositions align with the content. By contrast, liberal-leaning untrustworthy domains generally show either a smaller or a negative association. These heterogeneous patterns further underscore the ideological asymmetry found in our results.

Furthermore, the data for this study were collected in 2020. This timing provides a unique snapshot of belief formation during a critical period. The 2020 US presidential election and the “Stop the Steal” movement were foundational events that have had a lasting impact on the political landscape and the discourse around election integrity (Chen, Lukito, and Koo 2023; Starbird, DiResta, and DeButts 2023; Prochaska et al. 2023). Understanding the dynamics at play during this period is crucial for contextualizing the persistent narratives of election denialism that continue to shape politics. At the same time, we acknowledge the study’s limited temporal validity (Munger 2023), given that this study examines a unique outcome from a single election that occurred multiple electoral cycles ago and is limited to the United States (Lukito 2024). Said another way, the mechanisms of selective exposure and motivated reasoning we identify are possibly stable, but the specific content and platforms may change. Therefore, while our study offers insights into a critical event, we caution against direct extrapolation of the effect sizes to subsequent elections without further research.

Our data are limited in their representativeness. The data were collected from people in the United States who installed the YouGov Pulse plugin. This self-selection can impact the generalizability of the results (Gil-López et al. 2023; Hopkins and Gorton 2024). The YouGov Pulse panel, although high quality, is known to underrepresent individuals without a high school diploma and overrepresent Democrats (Guess, Nyhan, and Reifler 2020). However, the panel closely resembles the American adult population along many other dimensions (Guess, Nyhan, and Reifler 2020). There may also be concerns that participants modified their browsing behavior because they knew they were being observed (i.e., a Hawthorne effect (Saha et al. 2024)). However, recent research on this issue in the context of social media data donation finds that while some behavioral changes occur (e.g., a reduction in self-focused content) immediately after initial participation, these effects often diminish over time as participants habituate to being observed (Saha et al. 2024). Given that our data collection occurred over an extended period, and that YouGov Pulse participants are standing members of the panel, it is plausible that any initial observer effects would have attenuated. In addition, the prevalence of visits to sensitive websites in our data suggests that significant self-censorship was unlikely to have occurred. While we use survey weights provided by YouGov to improve representativeness when appropriate in the regression analyses, we acknowledge that the core relationships we identify are based on a sample that is not a true probability sample of the US population. We also only examine web-browsing data and not other spaces where untrustworthy content may be consumed, such as in personal messaging applications (Dahlke and Hancock 2025).

Finally, as an observational study employing double machine learning (Wager and Athey 2018), our estimation relies on the assumption of conditional ignorability: that no unobserved factors simultaneously confound both exposure and belief formation after conditioning on our extensive set of covariates. Although our methodology relaxes some parametric assumptions, this core requirement remains. Recognizing that this core assumption is untestable and the potential for omitted variable bias, we conducted rigorous sensitivity analyses to assess the robustness of our findings (see Appendix F). First, we calculate the Robustness Value (RV), representing the minimum percentage of residual variance in *both* exposure and belief that an unobserved confounder would need to nullify the association (Cinelli and Hazlett 2020). For the overall sample, this threshold is 6% (RV = 0.06, Table 5), while for the Trump supporter subgroup, the threshold is substantially higher at 11% (RV = 0.11, Table 5). Benchmarking these thresholds against the individual confounding potential of our observed covariates (Table 6) illustrates their stringency; for an *unmeasured* variable to explain away even the overall association, it would need to possess a stronger joint association with both exposure and belief than key predictors like prior untrustworthy website exposure or political knowledge demonstrate individually. Moreover, the higher RV for the theoretically predicted subgroup (Trump supporters) provides additional evidence for the robustness of the observed heterogeneity. As complementary checks, we performed the Oster test (Oster 2019) and placebo tests against other political behaviors. The Oster test (Table 7), which relies on linear assumptions and is known for its conservatism (failing in nearly half of studies in top economics journals (Oster 2019)), shows our association remains directionally positive even after its stringent adjustment; the small magnitude of the adjusted coefficient is characteristic of this demanding test and should be interpreted in light of the directional stability. Crucially, we also tested for associations between untrustworthy website exposure and a range of other related political activities (e.g., donating, volunteering, rally attendance; see Table 8). If our primary finding were merely an artifact of a general latent characteristic driving *both* exposure and political activism, such a confounder would likely induce spurious correlations with these other behaviors as well. The consistent lack of significant associations across these related outcomes challenges explanations relying on pervasive unobserved factors. Collectively, this multipronged sensitivity analysis—emphasizing the demanding R-value thresholds relative to observed benchmarks, the directional stability under the conservative Oster test, and the null results from placebo tests—provides evidence that the observed association, particularly its heterogeneity, is unlikely to be solely an artifact of plausible unobserved confounding.

References

Adams, Zoë, Magda Osman, Christos Bechlivanidis, and Björn Meder. 2023. "(Why) Is Misinformation a Problem?" *Perspectives on Psychological Science* 18, no. 6 (November 1, 2023): 1436–63. https://doi.org/10.1177/17456916221141344.

Ahmad, Amer Khaled, Abd Allah AL-Jalabneh, Alia Mahmoud, and Amjad Safori. 2022. "Covid-19 and the resurgence of the hypodermic needle theory applicability in times of crises." In *International Conference on Business and Technology,* 1423–36. Springer. https://doi.org/10.1007/978-3-031-08954-1_124.

Ahmad, Wajeeha, Ananya Sen, Charles Eesley, and Erik Brynjolfsson. 2024. "Companies Inadvertently Fund Online Misinformation Despite Consumer Backlash." *Nature* 630, no. 8015 (June 12, 2024): 123–31. https://doi.org/10.1038/s41586-024-07404-1.

Allcott, Hunt, and Matthew Gentzkow. 2017. "Social Media and Fake News in the 2016 Election." *Journal of Economic Perspectives* 31, no. 2 (May 1, 2017): 211–36. https://doi.org/10.1257/jep.31.2.211.

Allcott, Hunt, Matthew Gentzkow, Ro'ee Levy, Adriana Crespo-Tenorio, Natasha Dumas, Winter Mason, Devra Moehler, et al. 2025. *The Effects of Political Advertising on Facebook and Instagram Before the 2020 US Election.* NBER Working Paper 33818. National Bureau of Economic Research, May 31, 2025. https://doi.org/10.3386/w33818.

Allcott, Hunt, Matthew Gentzkow, Winter Mason, Arjun Wilkins, Pablo Barberá, Taylor Brown, Juan Carlos Cisneros, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, et al. 2024. "The Effects of Facebook and Instagram on the 2020 Election: A Deactivation Experiment." *Proceedings of the National Academy of Sciences* 121, no. 21 (May 13, 2024): e2321584121. https://doi.org/10.1073/pnas.2321584121.

Allen, Jennifer, Baird Howland, Markus Mobius, David Rothschild, and Duncan J. Watts. 2020. "Evaluating the Fake News Problem at the Scale of the Information Ecosystem." *Science Advances* 6, no. 14 (April 3, 2020): eaay3539. https://doi.org/10.1126/sciadv.aay3539.

Allen, Jennifer, Duncan J. Watts, and David G. Rand. 2024. "Quantifying the Impact of Misinformation and Vaccine-Skeptical Content on Facebook." *Science* 384, no. 6653 (May 10, 2024): eadk3451. https://doi.org/10.1126/science.eadk3451.

Altay, Sacha, Emma Hoes, and Magdalena Wojcieszak. 2025. "Following news on social media boosts knowledge, belief accuracy and trust." *Nature Human Behaviour,* 1–10. https://doi.org/10.1038/s41562-025-02205-6.

Altay, Sacha, Benjamin A. Lyons, and Ariana Modirrousta-Galian. 2024. "Exposure to Higher Rates of False News Erodes Media Trust and Fuels Overconfidence." *Mass Communication and Society* 27, no. 6 (August 7, 2024): 1–25. https://doi.org/10.1080/15205436.2024.2382776.

Altay, Sacha, Rasmus Kleis Nielsen, and Richard Fletcher. 2022. "Quantifying the 'Infodemic': People Turned to Trustworthy News Outlets During the 2020 Coronavirus Pandemic." *Journal of Quantitative Description: Digital Media* 2 (July 1, 2022): 1–24. https://doi.org/10.51685/jqd.2022.020.

Anderson, Ashley A, Dominique Brossard, Dietram A Scheufele, Michael A Xenos, and Peter Ladwig. 2014. "The "Nasty Effect:" Online Incivility And Risk Perceptions Of Emerging Technologies." *Journal of computer-mediated communication* 19 (3): 373–87. https://doi.org/10.1111/jcc4.12009.

Arceneaux, Kevin, Can Zengin, and Jonathan M. Ladd. n.d. "Stop Blaming Social Media for Everything: The Minimal Effects of Facebook in the 2024 US Presidential Election," accessed June 10, 2025. https://www.jonathanmladd.com/uploads/5/3/6/6/5366295/facebook_field_experiment_2024_u_s_.pdf.

Asker, David, and Elias Dinas. 2017. "Do Online Media Polarize? Evidence From The Comments' Section." *Evidence from the Comments' Section (August 11, 2017),* http://dx.doi.org/10.2139/ssrn.3018223.

Aslett, Kevin, Andrew M. Guess, Richard Bonneau, Jonathan Nagler, and Joshua A. Tucker. 2022. "News Credibility Labels Have Limited Average Effects on News Diet Quality and Fail to Reduce Misperceptions." *Science Advances* 8, no. 18 (May 6, 2022): eabl3844. https://doi.org/10.1126/sciadv.eabl3844.

Bak-Coleman, Joseph B., Stephan Lewandowsky, Philipp Lorenz-Spreen, Arvind Narayanan, Amy Orben, and Lisa Oswald. 2025. *Moving Towards Informative and Actionable Social Media Research.* Preprint: arXiv, May 14, 2025. https://doi.org/10.48550/arXiv.2505.09254.

Baron, Jonathan, and John T. Jost. 2019. "False Equivalence: Are Liberals and Conservatives in the United States Equally Biased?" *Perspectives on Psychological Science* 14, no. 2 (March 1, 2019): 292–303. https://doi.org/10.1177/1745691618788876.

Bello-Pardo, E.D. 2022. "A Digital Hypodermic Needle? Essays on the Impact of Misinformation, Framing, and Images on American Public Opinion in the Internet Age." PhD diss., American University. https://aura.american.edu/articles/thesis/A_Digital_Hypodermic_Needle_Essays_On_The_Impact_Of_Misinformation_Framing_And_Images_On_American_Public_Opinion_In_The_Internet_Age/23949411.

Bennett, W. Lance, and Shanto Iyengar. 2008. "A New Era of Minimal Effects? The Changing Foundations of Political Communication." *Journal of Communication* 58, no. 4 (December 29, 2008): 707–31. https://doi.org/10.1111/j.1460-2466.2008.00410.x.

Bineham, Jeffery L. 1988. "A historical account of the hypodermic model in mass communication." *Communications Monographs* 55 (3): 230–46.

Budak, Ceren, Brendan Nyhan, David M. Rothschild, Emily Thorson, and Duncan J. Watts. 2024. "Misunderstanding the Harms of Online Misinformation." *Nature* 630, no. 8015 (June 5, 2024): 45–53. https://doi.org/10.1038/s41586-024-07417-w.

Calvillo, Dustin P., and Thomas J. Smelter. 2020. "An Initial Accuracy Focus Reduces the Effect of Prior Exposure on Perceived Accuracy of News Headlines." *Cognitive Research: Principles and Implications* 5 (November 5, 2020): 55. https://doi.org/10.1186/s41235-020-00257-y.

Center for an Informed Public. n.d. "Conspiracy Theory Spreading on Social Media About Starlink Interfering with Election Results." Rapid Research Blog, Center for an Informed Public, University of Washington. Accessed May 4, 2025. https://www.cip.uw.edu/2024/11/18/conspiracy-theory-starlink-election-results/.

Chen, Bin, Josephine Lukito, and Gyo Hyun Koo. 2023. "Comparing the #StopTheSteal Movement Across Multiple Platforms: Differentiating Discourse on Facebook, Twitter, and Parler." *Social Media + Society* 9, no. 3 (September 19, 2023): 20563051231196879. https://doi.org/10.1177/20563051231196879.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. 2018. "Double/Debiased Machine Learning for Treatment and Structural Parameters." *The Econometrics Journal* 21, no. 1 (February 1, 2018): C1–C68. https://doi.org/10.1111/ectj.12097.

Cinelli, Carlos, and Chad Hazlett. 2020. "Making Sense of Sensitivity: Extending Omitted Variable Bias." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 82, no. 1 (February 12, 2020): 39–67. https://doi.org/10.1111/rssb.12348.

Craig, Stephen C., and Jason Gainous. 2024. "To Vote or Not to Vote? Fake News, Voter Fraud, and Support for Postponing the 2020 US Presidential Election." *Politics & Policy* 52, no. 1 (January 1, 2024): 33–50. https://doi.org/10.1111/polp.12577.

Dahlke, Ross, and Jeffrey Hancock. 2022. "The Effect of Online Misinformation Exposure on False Election Beliefs." *OSF Preprints* 17 (January 1, 2022). https://doi.org/10.31219/osf.io/325tn.

———. 2025. "The Public Sphere in Private Spaces: Quantifying Political Computer-Mediated Communication in Personal Messaging," https://doi.org/10.31219/osf.io/2fx3t_v1.

Dahlke, Ross, Deepak Kumar, Zakir Durumeric, and Jeffrey T. Hancock. 2023. "Quantifying the Systematic Bias in the Accessibility and Inaccessibility of Web Scraping Content from url-Logged Web-Browsing Digital Trace Data." *Social Science Computer Review* (November 29, 2023): 08944393231218214. https://doi.org/10.1177/08944393231218214.

Dahlke, Ross, and Jennifer Pan. 2024. "January 6 arrests and media coverage do not remobilize conservatives on social media." *Proceedings of the National Academy of Sciences* 121 (23): e2401239121. https://doi.org/10.1073/pnas.2401239121.

Ditto, Peter H., Cory J. Clark, Brittany S. Liu, Sean P. Wojcik, Eric E. Chen, Rebecca H. Grady, Jared B. Celniker, and Joanne F. Zinger. 2019. "Partisan Bias and Its Discontents." *Perspectives on Psychological Science* 14, no. 2 (March 1, 2019): 304–16. https://doi.org/10.1177/1745691618817753.

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. 2019. "At Least Bias Is Bipartisan: A Meta-Analytic Comparison of Partisan Bias in Liberals and Conservatives." *Perspectives on Psychological Science* 14, no. 2 (March 1, 2019): 273–91. https://doi.org/10.1177/1745691617746796.

Douglas, Karen M., Joseph E. Uscinski, Robbie M. Sutton, Aleksandra Cichocka, Turkay Nefes, Chee Siang Ang, and Farzin Deravi. 2019. "Understanding Conspiracy Theories." *Political Psychology* 40, no. 1 (February 1, 2019): 3–35. https://doi.org/10.1111/pops.12568.

Druckman, James N., and Mary C. McGrath. 2019. "The Evidence for Motivated Reasoning in Climate Change Preference Formation." *Nature Climate Change* 9, no. 2 (January 21, 2019): 111–19. https://doi.org/10.1038/s41558-018-0360-1.

Drummond, Caitlin, Michael Siegrist, and Joseph Árvai. 2020. "Limited effects of exposure to fake news about climate change." *Environmental Research Communications* 2 (8): 081003. https://doi.org/10.1088/2515-7620/abae77.

DuGoff, Eva H., Megan Schuler, and Elizabeth A. Stuart. 2014. "Generalizing Observational Study Results: Applying Propensity Score Methods to Complex Surveys." *Health Services Research* 49 (1): 284–303. https://doi.org/10.1111/1475-6773.12090.

Ecker, Ullrich, Jon Roozenbeek, Sander Van Der Linden, Li Qian Tay, John Cook, Naomi Oreskes, and Stephan Lewandowsky. 2024. "Misinformation Poses a Bigger Threat to Democracy Than You Might Think." *Nature* 630 (8015): 29–32. https://doi.org/10.1038/d41586-024-01587-3.

Epley, Nicholas, and Thomas Gilovich. 2016. "The Mechanics of Motivated Reasoning." *Journal of Economic Perspectives* 30, no. 3 (August 1, 2016): 133–40. https://doi.org/10.1257/jep.30.3.133.

Ernst, Nicole, Rinaldo Kühne, and Werner Wirth. 2017. "Effects of Message Repetition and Negativity on Credibility Judgments and Political Attitudes." *International Journal of Communication* 11 (February 1, 2017): 21. https://ijoc.org/index.php/ijoc/article/view/5998.

Fahey, James J. 2023. "The big lie: Expressive responding and misperceptions in the United States." *Journal of Experimental Political Science* 10 (2): 267–78. https://doi.org/10.1017/XPS.2022.33.

Fazio, Lisa K., Nadia M. Brashier, B. Keith Payne, and Elizabeth J. Marsh. 2015. "Knowledge Does Not Protect Against Illusory Truth." *Journal of Experimental Psychology: General* 144, no. 5 (October 1, 2015): 993–1002. https://doi.org/10.1037/xge0000098.

Featherstone, Jack D., and Jian Zhang. 2020. "Feeling Angry: The Effects of Vaccine Misinformation and Refutational Messages on Negative Emotions and Vaccination Attitude." *Journal of Health Communication* 25, no. 9 (October 4, 2020): 692–702. https://doi.org/10.1080/10810730.2020.1803978.

Federico, Christopher M, Grace Deason, and Emily L Fisher. 2012. "Ideological asymmetry in the relationship between epistemic motivation and political attitudes." *Journal of Personality and Social Psychology* 103 (3): 381. https://doi.org/10.1037/a0029063.

Freelon, Deen, and Chris Wells. 2020. "Disinformation as Political Communication." *Political Communication* 37, no. 2 (March 3, 2020): 145–56. https://doi.org/10.1080/10584609.2020.1723755.

Frenda, Shayna J., E. Derek Knowles, William Saletan, and Elizabeth F. Loftus. 2013. "False Memories of Fabricated Political Events." *Journal of Experimental Social Psychology* 49, no. 3 (May 1, 2013): 280–86. https://doi.org/10.1016/j.jesp.2012.12.001.

Garrett, R. Kelly, and Robert M. Bond. 2021. "Conservatives' Susceptibility to Political Misperceptions." *Science Advances* 7, no. 23 (June 1, 2021): eabf1234. https://doi.org/10.1126/sciadv.abf1234.

Gelbach, Jonah B. 2016. "When Do Covariates Matter? And Which Ones, and How Much?" *Journal of Labor Economics* 34, no. 3 (July 1, 2016): 509–43. https://doi.org/10.1086/683668.

Gil-López, Teresa, Clara Christner, Ernesto de León, Mykola Makhortykh, Aleksandra Urman, Michaela Maier, and Silke Adam. 2023. "Do (Not!) Track Me: Relationship Between Willingness to Participate and Sample Composition in Online Information Behavior Tracking Research." *Social Science Computer Review* 41, no. 6 (February 12, 2023): 2274–92. https://doi.org/10.1177/08944393231156634.

Goel, Pranav, Jon Green, David Lazer, and Philip S Resnik. 2025. "Using Co-Sharing to Identify Sse of Mainstream News for Promoting Potentially Misleading Narratives." *Nature Human Behaviour,* 1–18. https://doi.org/10.1038/s41562-025-02223-4.

Graham, Jesse, Jonathan Haidt, and Brian A. Nosek. 2009. "Liberals and Conservatives Rely on Different Sets of Moral Foundations." *Journal of Personality and Social Psychology* 96, no. 5 (May 1, 2009): 1029–46. https://doi.org/10.1037/a0015141.

Graham, Matthew H., and Omer Yair. 2023. "Expressive Responding and Belief in 2020 Election Fraud." *Political Behavior* 46, no. 2 (July 26, 2023): 1349–74. https://doi.org/10.1007/s11109-023-09875-w.

Green, Jon, William Hobbs, Stefan McCabe, and David Lazer. 2022. "Online Engagement with 2020 Election Misinformation and Turnout in the 2021 Georgia Runoff Election." *Proceedings of the National Academy of Sciences* 119, no. 34 (August 1, 2022): e2115900119. https://doi.org/10.1073/pnas.2115900119.

Greene, Ciara M., Constance de Saint Laurent, Karen Hegarty, and Gillian Murphy. 2022. "False Memories for True and False Vaccination Information Form in Line with Pre-Existing Vaccine Opinions." *Applied Cognitive Psychology* 36, no. 6 (October 4, 2022): 1200–1208. https://doi.org/10.1002/acp.4002.

Grinberg, Nir, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. "Fake News on Twitter During the 2016 US Presidential Election." *Science* 363, no. 6425 (January 25, 2019): 374–78. https://doi.org/10.1126/science.aau2706.

Guess, Andrew, Jonathan Nagler, and Joshua A. Tucker. 2019. "Less Than You Think: Prevalence and Predictors of Fake News Dissemination on Facebook." *Science Advances* 5, no. 1 (January 1, 2019): eaau4586. https://doi.org/10.1126/sciadv.aau4586.

Guess, Andrew M., Pablo Barberá, Simon Munzert, and JungHwan Yang. 2021. "The Consequences of Online Partisan Media." *Proceedings of the National Academy of Sciences* 118, no. 14 (March 29, 2021): e2013464118. https://doi.org/10.1073/pnas.2013464118.

Guess, Andrew M., and Benjamin A. Lyons. 2020. "Misinformation, Disinformation, and Online Propaganda." In *Social Media and Democracy: The State of the Field, Prospects for Reform,* edited by Nathaniel Persily and Joshua A. Tucker, 10–33. Cambridge University Press, August 24, 2020. https://doi.org/10.1017/9781108890960.003.

Guess, Andrew M., Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow, et al. 2023a. "How Do Social Media Feed Algorithms Affect Attitudes and Behavior in an Election Campaign?" *Science* 381, no. 6656 (July 27, 2023): 398–404. https://doi.org/10.1126/science.abp9364.

———. 2023b. "Reshares on Social Media Amplify Political News but Do Not Detectably Affect Beliefs or Opinions." *Science* 381, no. 6656 (July 27, 2023): 404–8. https://doi.org/10.1126/science.add8424.

Guess, Andrew M., Brendan Nyhan, and Jason Reifler. 2020. "Exposure to Untrustworthy Websites in the 2016 US Election." *Nature Human Behaviour* 4, no. 5 (May 1, 2020): 472–80. https://doi.org/10.1038/s41562-020-0833-x.

Hassan, Aumyo, and Sarah J. Barber. 2021. "The Effects of Repetition Frequency on the Illusory Truth Effect." *Cognitive Research: Principles and Implications* 6, no. 1 (May 13, 2021): 38. https://doi.org/10.1186/s41235-021-00301-5.

Ho, Daniel, Kosuke Imai, Gary King, and Elizabeth A. Stuart. 2011. "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference." *Journal of Statistical Software* 42, no. 8 (September 27, 2011): 1–28. https://doi.org/10.18637/jss.v042.i08.

Hopkins, Daniel J., and Tori Gorton. 2024. "On the Internet, No One Knows You're an Activist: Patterns of Participation and Response in an Online, Opt-in Survey Panel." *Political Research Quarterly* 77, no. 4 (August 4, 2024): 1397–414. https://doi.org/10.1177/10659129241268818.

Jalbert, Madeline, Eryn Newman, and Norbert Schwarz. 2020. "Only Half of What I'll Tell You Is True: Expecting to Encounter Falsehoods Reduces Illusory Truth." *Journal of Applied Research in Memory and Cognition* 9, no. 4 (December 1, 2020): 602–13. https://doi.org/10.1016/j.jarmac.2020.06.001.

Jost, John T., Jack Glaser, Frank J. Sulloway, and Arie W. Kruglanski. 2018. "Political Conservatism as Motivated Social Cognition." In *The Motivated Mind,* 129–204. Routledge, May 1, 2018.

Justwan, Florian, Bert Baumgaertner, and Madeleine Curtright. 2024. "Meddling in the 2016 Elections and Satisfaction With Democracy in the US." *Political Studies* 72, no. 2 (May 1, 2024): 505–26. https://doi.org/10.1177/00323217221126300.

Kahan, Dan M. 2017. "Misconceptions, Misinformation, and the Logic of Identity-Protective Cognition." Cultural Cognition Project Working Paper No. 164, *SSRN Electronic Journal* (January 1, 2017). https://doi.org/10.2139/ssrn.2973067.

Kunda, Ziva. 1990. "The Case for Motivated Reasoning." *Psychological Bulletin* 108, no. 3 (November 1, 1990): 480–98. https://doi.org/10.1037/0033-2909.108.3.480.

Lavigne, Mathieu, Brian Fogarty, John Carey, Brendan Nyhan, and Jason Reifler. n.d. *Inattention and Differential Exposure: How Media Questioning of Election Fraud Misinformation Often Fails to Reach the Public.* Unpublished manuscript. Accessed June 10, 2025. https://business.columbia.edu/sites/default/files-efs/imce-uploads/nyhan-fraud-exposure-online.pdf.

Li, Fan, Kari Lock Morgan, and Alan M Zaslavsky. 2018. "Balancing covariates via propensity score weighting." *Journal of the American Statistical Association* 113 (521): 390–400. https://doi.org/10.1080/01621459.2016.1260466.

Lukito, Josephine. 2024. "Global Misinformation & Disinformation Special Issue Introduction." *International Journal of Public Opinion Research* 36, no. 3 (October 1, 2024): edae030. https://doi.org/10.1093/ijpor/edae030.

McGowan, Lisa D. 2022. "tipr: An R Package for Sensitivity Analyses for Unmeasured Confounders." *Journal of Open Source Software* 7, no. 66 (February 1, 2022): 4495. https://doi.org/10.21105/joss.04495.

Moore, Ryan C., Ross Dahlke, and Jeffrey T. Hancock. 2023. "Exposure to Untrustworthy Websites in the 2020 US Election." *Nature Human Behaviour* 7, no. 7 (April 13, 2023): 1096–105. https://doi.org/10.1038/s41562-023-01564-2.

Munger, Kevin. 2023. "Temporal Validity as Meta-Science." *Research & Politics* 10, no. 3 (September 25, 2023): 20531680231187271. https://doi.org/10.1177/20531680231187271.

Murphy, Grainne, Elizabeth F. Loftus, Rachel H. Grady, Lori J. Levine, and Claire M. Greene. 2019. "False Memories for Fake News During Ireland's Abortion Referendum." *Psychological Science* 30, no. 11 (November 1, 2019): 1449–59. https://doi.org/10.1177/0956797619867129.

Nadarevic, Lena, and Edgar Erdfelder. 2014. "Initial Judgment Task and Delay of the Final Validity-Rating Task Moderate the Truth Effect." *Consciousness and Cognition* 23 (January): 74–84. https://doi.org/10.1016/j.concog.2013.12.002.

Nelson, Jacob L., and Harsh Taneja. 2018. "The Small, Disloyal Fake News Audience: The Role of Audience Availability in Fake News Consumption." *New Media & Society* 20, no. 10 (February 18, 2018): 3720–37. https://doi.org/10.1177/1461444818758715.

Newman, Eryn J., Madeline C. Jalbert, Norbert Schwarz, and Deva P. Ly. 2020. "Truthiness, the Illusory Truth Effect, and the Role of Need for Cognition." *Consciousness and Cognition* 78 (September 1, 2020): 102866. https://doi.org/10.1016/j.concog.2020.102866.

Nyhan, Brendan. 2020. "Facts and Myths about Misperceptions." *Journal of Economic Perspectives* 34, no. 3 (August 1, 2020): 220–36. https://doi.org/10.1257/jep.34.3.220.

Nyhan, Brendan, and Jason Reifler. 2010. "When Corrections Fail: The Persistence of Political Misperceptions." *Political Behavior* 32, no. 3 (September 1, 2010): 303–30. https://doi.org/10.1007/s11109-010-9112-2.

Nyhan, Brendan, Jaime Settle, Emily Thorson, Magdalena Wojcieszak, Pablo Barberá, Annie Y. Chen, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, et al. 2023. "Like-Minded Sources on Facebook are Prevalent but Not Polarizing." *Nature* 620, no. 7972 (July 23, 2023): 137–44. https://doi.org/10.1038/s41586-023-06390-w.

Oster, Emily. 2019. "Unobservable Selection and Coefficient Stability: Theory and Evidence." *Journal of Business & Economic Statistics* 37, no. 2 (April 1, 2019): 187–204. https://doi.org/10.1080/07350015.2018.1445820.

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. https://doi.org/10.1037/xge0000465.

Pennycook, Gordon, and David G. Rand. 2019. "Lazy, Not Biased: Susceptibility to Partisan Fake News Is Better Explained by Lack of Reasoning Than by Motivated Reasoning." *Cognition* 188 (May 1, 2019): 39–50. https://doi.org/10.1016/j.cognition.2018.11.011.

Petersen, Michael Bang. 2023. *Background Report: Empowering Audiences – Against Misinformation Through 'Prebunking'.* Technical report. The Future of Free Speech, Vanderbilt University, January 1, 2023. Accessed June 10, 2025. https://futurefreespeech.org/background-report-empowering-audiences-against-misinformation-through-prebunking/.

Peterson, Erik, and Shanto Iyengar. 2021. "Partisan Gaps in Political Information and Information-Seeking Behavior: Motivated Reasoning or Cheerleading?" *American Journal of Political Science* 65, no. 1 (January 1, 2021): 133–47. https://doi.org/10.1111/ajps.12535.

Prochaska, Stephen, Kayla Duskin, Zarine Kharazian, Carly Minow, Stephanie Blucker, Sylvie Venuto, Jevin D. West, and Kate Starbird. 2023. "Mobilizing Manufactured Reality: How Participatory Disinformation Shaped Deep Stories to Catalyze Action During the 2020 U.S. Presidential Election." *Proceedings of the ACM on Human–Computer Interaction* 7, no. CSCW1 (April 16, 2023): 1–39. https://doi.org/10.1145/3579616.

Robertson, Ronald E. 2022. "Uncommon Yet Consequential Online Harms." *Journal of Online Trust and Safety* 1 (3). https://doi.org/10.54501/jots.v1i3.87.

Robertson, Ronald E., Shan Jiang, Kenneth Joseph, Lisa Friedland, David Lazer, and Christo Wilson. 2018. "Auditing Partisan Audience Bias Within Google Search." *Proceedings of the ACM on Human-Computer Interaction* 2, no. CSCW (November 1, 2018): 148:1–148:22. https://doi.org/10.1145/3274417.

Saha, Koustuv, Pranshu Gupta, Gloria Mark, Emre Kıcıman, and Munmun De Choudhury. 2024. "Observer Effect in Social Media Use." In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems,* 1–20. Association for Computing Machinery, May 11, 2024. https://doi.org/10.1145/3613904.3642078.

Saint Laurent, Celine de, Grainne Murphy, Kieran Hegarty, and Claire M. Greene. 2022. "Measuring the Effects of Misinformation Exposure and Beliefs on Behavioural Intentions: A COVID-19 Vaccination Study." *Cognitive Research: Principles and Implications* 7, no. 1 (December 1, 2022): 87. https://doi.org/10.1186/s41235-022-00441-4.

Schaub, Sergei, and ETH Zurich. 2020. "Robomit: robustness checks for omitted variable bias." *R package version* 1 (3).

Sears, David O., and Jonathan L. Freedman. 1967. "Selective Exposure to Information: A Critical Review." *Public Opinion Quarterly* 31, no. 2 (June 1, 1967): 194–213. https://doi.org/10.1086/267513.

Slater, Michael D. 2007. "Reinforcing Spirals: The Mutual Influence of Media Selectivity and Media Effects and Their Impact on Individual Behavior and Social Identity." *Communication Theory* 17, no. 3 (August 1, 2007): 281–303. https://doi.org/10.1111/j.1468-2885.2007.00303.x.

Starbird, Kate, Renée DiResta, and Matt DeButts. 2023. "Influence and Improvisation: Participatory Disinformation During the 2020 US Election." *Social Media + Society* 9, no. 2 (June 7, 2023): 20563051231177943. https://doi.org/10.1177/20563051231177943.

Stroud, Natalie Jomini. 2008. "Media Use and Political Predispositions: Revisiting the Concept of Selective Exposure." *Political Behavior* 30, no. 3 (September 21, 2008): 341–66. https://doi.org/10.1007/s11109-007-9050-9.

———. 2010. "Polarization and Partisan Selective Exposure." *Journal of Communication* 60, no. 3 (June 1, 2010): 556–76. https://doi.org/10.1111/j.1460-2466.2010.01497.x.

Taber, Charles S., and Milton Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50, no. 3 (July 1, 2006): 755–69. https://doi.org/10.1111/j.1540-5907.2006.00214.x.

Tibshirani, Julie, Susan Athey, Erik Sverdrup, and Stefan Wager. 2024. *grf: Generalized Random Forests.* R package version 2.4.0. November 15, 2024. https://doi.org/10.32614/CRAN.package.grf.

Toepfl, Florian, and Eunike Piwoni. 2015. "Public Spheres In Interaction: Comment Sections Of News Websites As Counterpublic Spaces." *Journal of Communication* 65 (3): 465–88. https://doi.org/10.1111/jcom.12156.

Udry, Jessica, and Sarah J. Barber. 2024. "The Illusory Truth Effect: A Review of How Repetition Increases Belief in Misinformation." *Current Opinion in Psychology* 56 (April 1, 2024): 101736. https://doi.org/10.1016/j.copsyc.2023.101736.

Van Bavel, Jay J., and Andrea Pereira. 2018. "The Partisan Brain: An Identity-Based Model of Political Belief." *Trends in Cognitive Sciences* 22, no. 3 (March 1, 2018): 213–24. https://doi.org/10.1016/j.tics.2018.01.004.

Vellani, Valentina, Sarah Zheng, Dilay Ercelik, and Tali Sharot. 2023. "The Illusory Truth Effect Leads to the Spread of Misinformation." *Cognition* 236 (October 1, 2023): 105421. https://doi.org/10.1016/j.cognition.2023.105421.

Ventura, Tiago, Rajeshwari Majumdar, Jonathan Nagler, and Joshua A. Tucker. 2023. "Misinformation Beyond Traditional Feeds: Evidence from a WhatsApp Deactivation Experiment in Brazil." *SSRN Electronic Journal* (May 23, 2023). https://doi.org/10.2139/ssrn.4457400.

Wager, Stefan, and Susan Athey. 2018. "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests." *Journal of the American Statistical Association* 113, no. 523 (June 6, 2018): 1228–42. https://doi.org/10.1080/01621459.2017.1319839.

Weeks, Benjamin E., and Homero Gil de Zúñiga. 2021. "What's Next? Six Observations for the Future of Political Misinformation Research." *American Behavioral Scientist* 65, no. 2 (February 1, 2021): 277–89. https://doi.org/10.1177/0002764220905535.

World Economic Forum. 2024. *Global Risks Report 2024.* World Economic Forum Report, January 17, 2024. Accessed June 10, 2025. https://www.weforum.org/reports/global-risks-report-2024.

Zanutto, Elaine L. 2006. "A Comparison of Propensity Score and Linear Regression Analysis of Complex Survey Data." *Journal of Data Science* 4 (1): 67–91. https://doi.org/10.6339/JDS.2006.04(1).233.

Zhou, Alvin, Tian Yang, and Sandra González-Bailón. 2025. "The puzzle of misinformation: Exposure to unreliable content in the United States is higher among the better informed." *new media & society* 27 (3): 1526–43. https://doi.org/10.1177/14614448231196863.

Zhou, Yang, and Li Shen. 2022. "Confirmation Bias and the Persistence of Misinformation on Climate Change." *Communication Research* 49, no. 4 (June 1, 2022): 500–523. https://doi.org/10.1177/0093650220957029.

Zillmann, Dolf, and Jennings Bryant. 2013. *Selective Exposure to Communication.* Routledge, July 4, 2013. ISBN: 9780415515757. https://www.routledge.com/Selective-Exposure-To-Communication/Zillmann-Bryant/p/book/9780415515757.

Authors

Ross Dahlke is an Assistant Professor at the University of Wisconsin-Madison School of Journalism and Mass Communication. He studies information exposure on the open web and in personal messaging channels. His work has been published in Proceedings of the National Academy of Sciences (PNAS), Nature Human Behavior, Journal of Computer-Mediated Communication, and Proceedings of the ACM on Human-Computer Interaction (CSCW).

Jeffrey T. Hancock is the founding director of the Stanford Social Media Lab and is Harry and Norman Chandler Professor of Communication at Stanford University.

Acknowledgments

The authors thank participants of the UC Berkley School of Information 2022 Distinguished Lecture Series: Trustworthy Information; the 2023 Polarization Research Lab annual meeting; the 2023 Trust and Safety Research Conference; the 2023 Stanford Data Science Science Conference; members of the Stanford Social Media Lab, including Cid Decatur, Bingxu Han, Angela Lee, Sunny Liu, Ryan C. Moore, Lindsay Popowski, Ronald Robertson, Aysu Sarigul, Will Schulz, Ilana Shumsky, Debbie Siegel, Anja Stevic, Fanging Tu, Yi-Chia Wang, Sarah Wu, and Harry Yaojun Yan; and members of the Media and Democracy Data Cooperative, including Porismita Borha, Megan Brown, Bin Chen, Jason Greenfield, Sang Jung Kim, Jo Lukito, Maggie Macdonald, Meredith Pruden, Sarah Shugars, Jiyoun Suk, Yini Zhang, Weigh Zhong, and Yunkang Yang, for feedback on this work.

Data availability statement

The data underlying this article are available in the Open Science Framework at https://osf.io/egnwd/ (Dahlke and Hancock 2022).

Funding statement

This project was funded by a MURI award no. W911NF-20-1-0252 and support from Stanford's Freeman Spogli Institute for International Studies. R.D. was supported by graduate fellowship awards from Knight-Hennessy Scholars and Stanford Data Science at Stanford University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Ethical standards

This study was approved by the Stanford University Institutional Review Board [protocol no. IRB-53941]. Informed consent was obtained from all of the participants, who also received incentives from YouGov, the survey company that collected the data.

Keywords

Digital trace data, double machine learning, data science, false beliefs, causal inference.

Appendices

Appendix A: Regression Analyses and Gelbach Decompositions

To provide a parametric comparison to our main results, we present a series of weighted OLS regression models. To enhance the generalizability of these descriptive estimates, all models are weighted using the survey weights provided by YouGov, which adjust the sample to more closely match key demographic and political benchmarks of the US adult population. We present the results as a Gelbach Decomposition (Gelbach 2016) in Table 1. Descriptively, we find that people exposed to untrustworthy websites were more likely to believe that the 2020 US presidential election was won by the candidate who was not the certified winner than those who were not. Those exposed were 17.3% (95% CI = [12.7, 21.9]) more likely to believe the election was not won by the certified winner. We decompose the 17.3% difference using a Gelbach Decomposition into two components: the part of the difference that can be explained by observed individual characteristics, such as demographics and presidential candidate support, and the unexplained portion, which may represent the ideological asymmetric differences of exposure on election beliefs. The results of the Gelbach Decomposition (Table 1) reveal that 65.7% of this difference between those exposed and not exposed to false election beliefs can be explained by observed characteristics, such as demographics and political support. Therefore, 34.2% of the difference remains unexplained and may, in part, represent the effect of untrustworthy website exposure on false election beliefs. Another way to interpret these results is that model (1) represents a regression with no control variables and model (10) is a regression with the full control variables. Thus, the untrustworthy website exposure coefficient is akin to the results one could report in a simple regression-based analysis.

Table 1: OLS regression of the relationship of untrustworthy website exposure with the false belief, consecutively adding more covariates

	Belief that Trump won the 2020 US Presidential Election									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Untrustworthy Website Exposures	0.173***	0.077***	0.079***	0.083***	0.083***	0.085***	0.093***	0.089***	0.082***	0.059**
	(0.023)	(0.020)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.021)	(0.023)
Constant	0.135***	-0.014	-0.013	-0.019	0.006	0.006	0.061	-0.094*	-0.022	-0.020
	(0.014)	(0.013)	(0.014)	(0.020)	(0.022)	(0.025)	(0.034)	(0.042)	(0.047)	(0.047)
Conservative	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Total Web Visits Wave 1	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Education	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Race	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Political Knowledge	No	No	No	No	No	No	Yes	Yes	Yes	Yes
Political Interest	No	No	No	No	No	No	No	Yes	Yes	Yes
Age	No	No	No	No	No	No	No	No	Yes	Yes
Untrustworthy Website Exposure Wave 1	No	No	No	No	No	No	No	No	No	Yes
Observations	1,194	1,194	1,194	1,194	1,194	1,194	1,194	1,194	1,194	1,194
R^2	0.044	0.336	0.337	0.340	0.343	0.346	0.349	0.368	0.378	0.382
Adjusted R^2	0.043	0.335	0.335	0.337	0.340	0.343	0.343	0.361	0.370	0.374
Residual Std. Error	0.393	0.327	0.327	0.327	0.326	0.326	0.325	0.321	0.318	0.318
F Statistic	54.485***	302.005***	201.214***	101.992***	88.627***	62.551***	57.656***	57.235***	47.808***	45.492***

Note: Gelbach Decomposition of the difference of untrustworthy website exposure on the belief that Donald Trump won the 2020 US presidential election. Independent variables are sequentially added to the baseline model up to the final model. This decomposition reveals that about 70% of the difference in the belief that Trump won the election between those who were exposed to untrustworthy websites and those who were not can be explained by observed differences. About 30% of the difference remains unexplained and may partially be due to the effect of untrustworthy website exposure. Significance level: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

We repeat this analysis with dosage as the independent variable (Table 2).

Table 2: OLS regression of the dosage relationship of untrustworthy website exposure with the false belief, consecutively adding more covariates

	Belief that Trump won the 2020 US presidential election									
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Untrustworthy Website Exposures	0.001***	0.0005***	0.0005***	0.001***	0.001***	0.001***	0.001***	0.0005***	0.0004***	0.0004**
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Constant	0.188***	0.005	0.001	-0.004	0.018	0.021	0.067*	-0.087*	-0.018	-0.016
	(0.012)	(0.012)	(0.014)	(0.019)	(0.018)	(0.021)	(0.025)	(0.034)	(0.042)	(0.047)
Conservative	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Total Web Visits Wave 1	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Education	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Gender	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Race	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes
Political Knowledge	No	No	No	No	No	No	Yes	Yes	Yes	Yes
Political Interest	No	No	No	No	No	No	No	Yes	Yes	Yes
Age	No	No	No	No	No	No	No	No	Yes	Yes
Untrustworthy Website Exposure Wave 1	No	No	No	No	No	No	No	No	No	Yes
Observations	1,194	1,194	1,194	1,194	1,194	1,194	1,194	1,194	1,194	1,194
R^2	0.025	0.338	0.338	0.341	0.344	0.346	0.348	0.367	0.378	0.384
Adjusted R^2	0.024	0.336	0.336	0.338	0.341	0.342	0.360	0.370	0.375	0.375
Residual Std. Error	0.396	0.327	0.327	0.326	0.326	0.326	0.325	0.321	0.319	0.317
F Statistic	30.334***	303.499***	202.325***	102.540***	88.889***	62.630***	57.465***	56.978***	47.650***	45.810***

Note: Gelbach Decomposition of the difference of untrustworthy website exposure on the belief that Donald Trump won the 2020 US presidential election. Independent variables are sequentially added to the baseline model up to the final model. This decomposition reveals that about 50% of the difference in the belief that Trump won the election between those who were exposed to untrustworthy websites and those who were not can be explained by observed differences. About 50% of the difference remains unexplained and may partially be due to the effect of untrustworthy website exposure. Significance level: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Appendix B: Model Calibration and Robustness

We examine the calibration and robustness of the models in two main ways: calibration testing and estimating the associations using an alternative double machine learning methodology. First, the calibration tests show that two models (predicting exposure and predicting differences in outcome) used for assessing the association between untrustworthy website exposure and the belief that the election was fraudulent are well-calibrated. The calibration test is done as "best linear fit using forest predictions (on held-out data) as well as the mean forest prediction as regressors, along with one-sided heteroskedasticity-robust (HC3) SEs" (Tibshirani et al. 2024), implemented directly in the `grf` package. The first model is well-calibrated and accurately predicts exposure ($p < .001$; AUC = 0.754; see Figure 3). The second model is also well-calibrated and accurately predicts belief ($p < .001$; AUC = 0.752; see Figure 4).

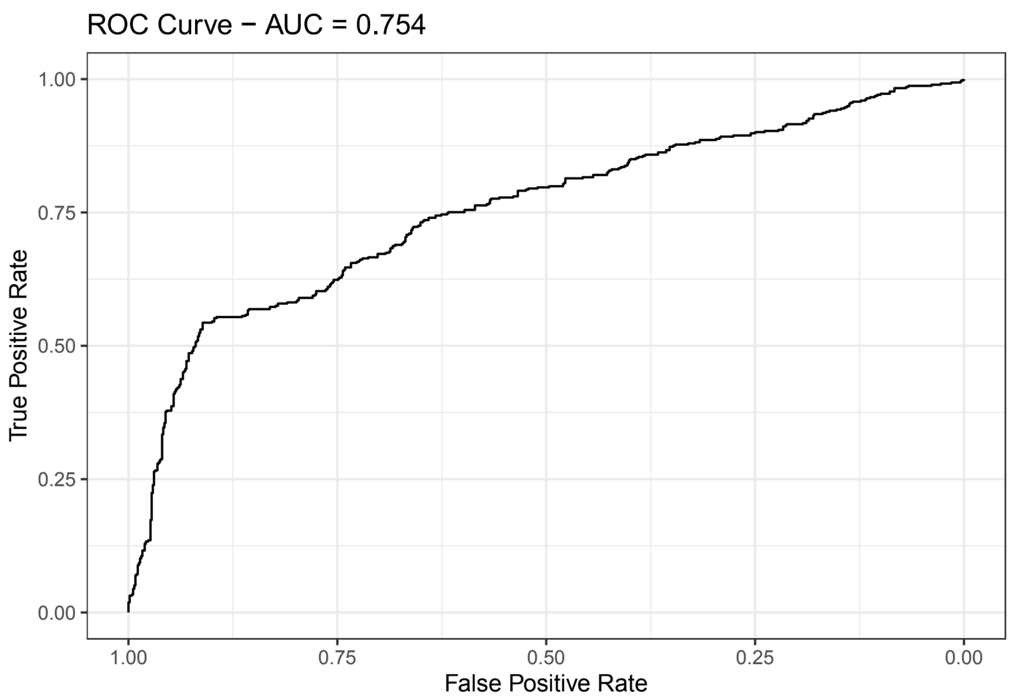

Figure 3: Receiver Operating Characteristic (ROC) curve for the predictive model of untrustworthy website exposure. The area under the curve (AUC) of 0.754 indicates a good level of accuracy in the model's ability to distinguish between individuals who were and were not exposed to untrustworthy websites. The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1 - specificity) at various threshold settings. The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test. The point where the curve touches the top border represents the threshold with the best balance between sensitivity and specificity. The calibration of the model was assessed using a best linear fit on held-out data with the inclusion of mean forest predictions as regressors, and the statistical significance was established using one-sided heteroskedasticity-robust (HC3) standard errors.

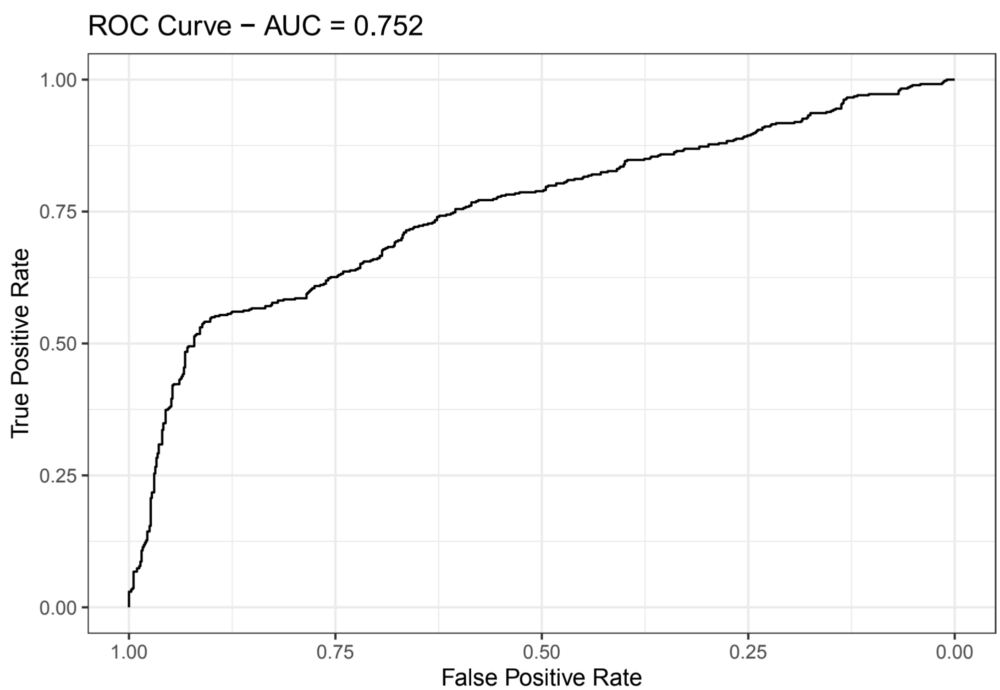

Figure 4: Receiver Operating Characteristic (ROC) curve for the predictive model of the belief that the 2020 US presidential election was fraudulent. The area under the curve (AUC) of 0.752 indicates a good level of accuracy in the model's ability to distinguish between individuals who were and were not exposed to untrustworthy websites. The ROC curve plots the true positive rate (sensitivity) against the false positive rate (1 - specificity) at various threshold settings. The closer the curve follows the left-hand border and then the top border of the ROC space, the more accurate the test. The point where the curve touches the top border represents the threshold with the best balance between sensitivity and specificity. The calibration of the model was assessed using a best linear fit on held-out data with the inclusion of mean forest predictions as regressors, and the statistical significance was established using one-sided heteroskedasticity-robust (HC3) standard errors.

Furthermore, we calculate results under related methodologies, finding that the results remain robust to these additional analyses. First, we calculate results using Chernozhukov's (2018) double machine learning method. Using this alternative method, we find the results remain significant, with an average association of 5.7% (95% CI = [1.4, 10.1]) for untrustworthy website exposure and 0.035% (95% CI = [0.023, 0.048]) for each additional exposure. Using a propensity matching model (Ho et al. 2011), the results are also significant, with an average difference of 6.6% (95% CI = [2.4, 10.9], $p = 0.002$) for untrustworthy website exposure and 0.037 (95% CI = [0.022, 0.051], $p < 0.001$) for the dosage.

Appendix C: Model Explanation

To provide an additional understanding of the machine learning models, we provide underlying statistical results used in the ultimate calculations of the conditional average association. First, we provide variable importance information from the first causal forest model predicting exposure (Figure 5). Variable importance is a measure of the relative contribution of each predictor variable to the model's predictive power. It is derived from the model's algorithm, indicating how much the model's accuracy decreases when data for that variable is permuted. In the context of a causal forest, this measure helps identify which variables are most predictive of treatment assignment—in this case, exposure to untrustworthy websites. High variable importance suggests that the variable plays a significant role in the model's decisions, whereas low importance indicates a smaller role. The visualization in Figure 5 presents these importance scores, ranking the variables from the most to the least important in predicting untrustworthy website exposure. Variables such as prior untrustworthy website exposure, political knowledge, and conservative political alignment appear to be strong predictors, highlighting the role of preexisting beliefs and information consumption patterns in determining exposure to untrustworthy websites.

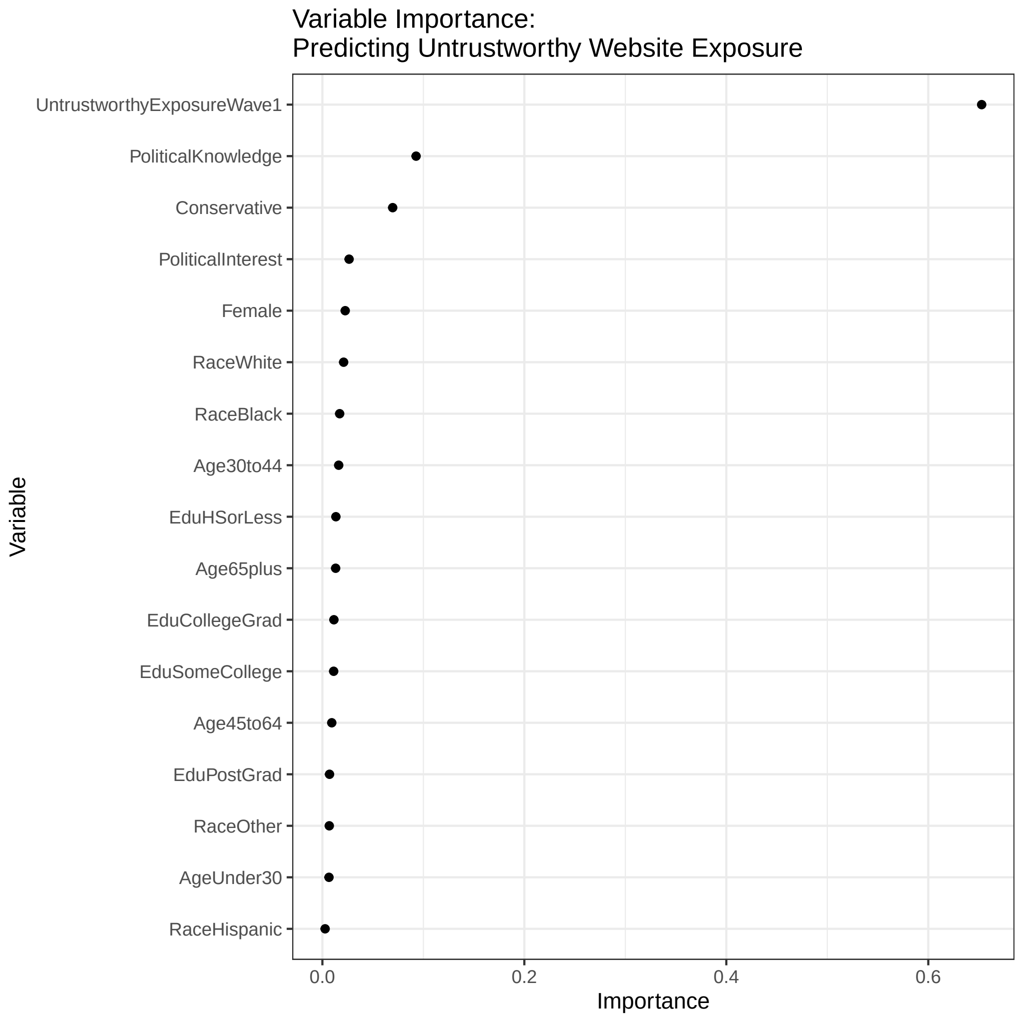

Figure 5: Variable importance scores as determined by the double machine learning model for predicting exposure to untrustworthy websites. These scores quantify the contribution of each predictor variable to the model's predictive accuracy, with higher values indicating greater importance. The determination of variable importance is based on the mean decrease in accuracy when the variable's values are permuted across the out-of-bag samples, a common method in random forest algorithms. This metric reflects the relative significance of each variable within the model; for example, "UntrustworthyWebsiteExposureWave1" shows the highest importance, suggesting it is the most critical predictor in the model.

Furthermore, to better show the underlying predictions of this first model, we present Figure 6. This figure shows the distribution of the predicted propensity to be treated or the number of predicted exposures. Again, this figure shows that the models are generally accurate in predicting exposure to untrustworthy websites.

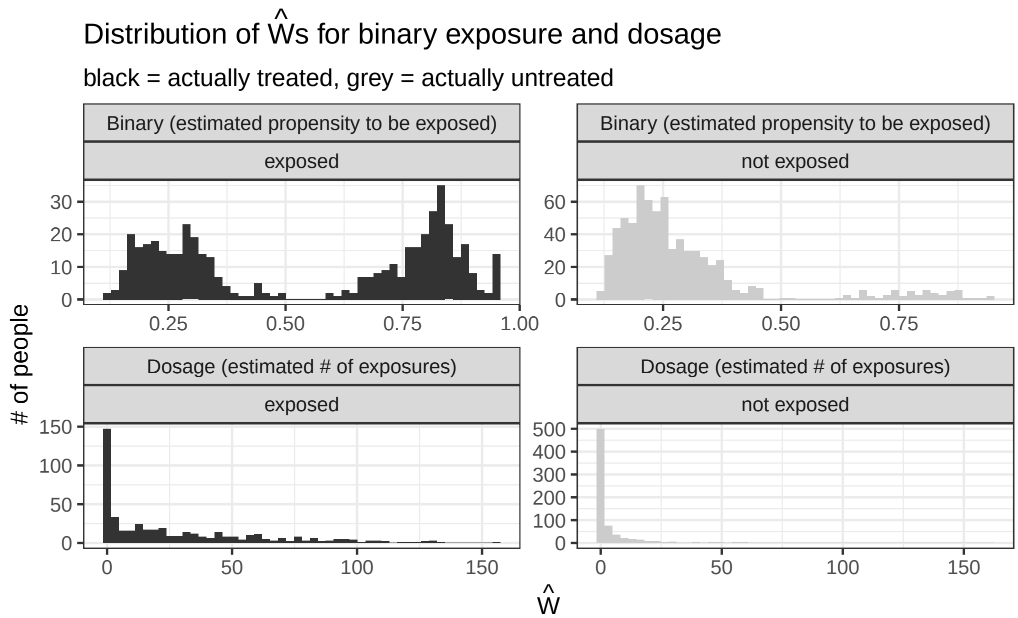

Figure 6: Distribution of propensities for binary exposure and estimated number of untrustworthy websites an individual was exposed to. On the x-axis is this estimated number. In the top panel, the x-axis is the estimated propensity to be treated—the likelihood of being exposed to an untrustworthy website. In the bottom panel, the x-axis is the expected number of untrustworthy websites each individual is exposed to. The y-axis for both panels is the count of people. The black bars represent people who were actually treated (exposed to at least one untrustworthy website), and the grey bars represent people who were not exposed.

Next, we show how these propensities are incorporated into the second machine learning model output by showing the propensity to be exposed, the calculated individual associations (ITEs), and the model weight in Figure 7.

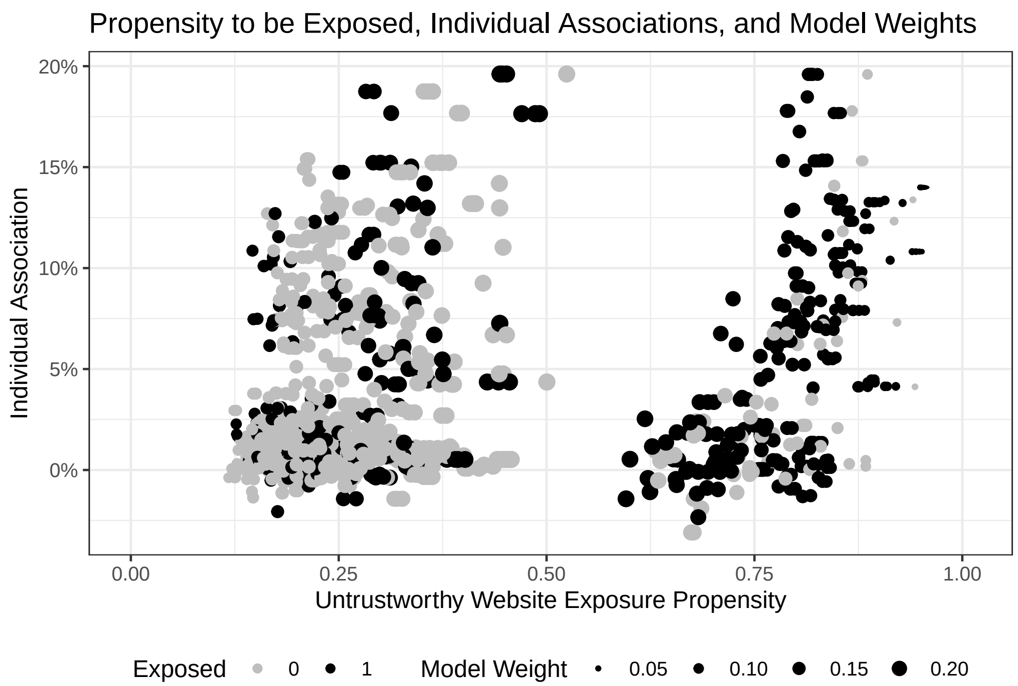

Figure 7: Plot of the propensity of being exposed to an untrustworthy website and the individual associations (ITEs). Points are sized by the model weight (based on the propensity to be exposed) and are colored by whether the individual was actually treated (exposed to an untrustworthy website).

Appendix D: Individual Associations

Under the expected outcomes framework, we calculate each individual's individual association and show conservatives' individual associations in a caterpillar plot in Figure 8.

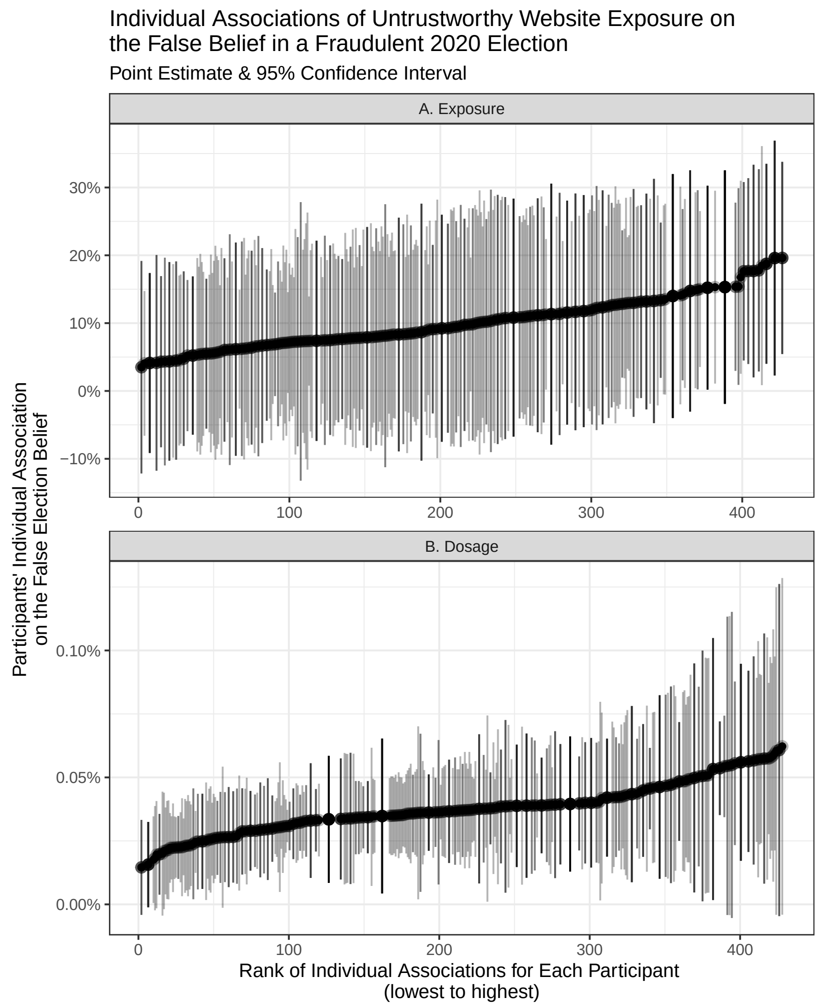

Figure 8: Plot of estimated individual associations and 95% confidence intervals of untrustworthy website exposure on the belief that Donald Trump won the 2020 US presidential election for each conservative individual in our sample. Participants are ordered along the x-axis from the lowest conditional estimate to the highest, and the y-axis shows the estimated individual association.

Appendix E: Heterogeneous Results by Website Ideology

To examine whether the relationship between untrustworthy website exposure and false election beliefs varies by the ideological slant of the visited domains, we further stratified our untrustworthy website measures using domain ratings from Robertson et al. (2018). These ratings estimate the ideology of the domain on a scale from -1 to 1, with -1 being the most liberal and 1 being the most conservative. Domains that were visited by our participants with a rating less than 0 were coded as *liberal* and those with a rating greater than 0 as *conservative*. We then re-estimated our double machine learning models separately for each subset, for both the binary exposure and dosage measures. To test the robustness of this finding, we also examined alternative cutoff points ranging from ±0.1 to ±0.8 (see Tables 3 and 4), finding results generally hold across cutoffs.

Conservative Domains. For the binary exposure measure for conservative domains, the overall estimated association was 5.5% (95% CI [1.0, 9.9]). In subgroup analyses, the conditional average association for conservative individuals and the false election belief was 13.8% (95% CI [2.6, 24.9]), whereas among liberals the difference was negligible (0.2%; 95% CI [–1.8, 2.2]). For the dosage measure, the overall dosage association was 0.035% (95% CI [0.017, 0.054]).

Liberal Domains. For the binary exposure measure for liberal domains, the overall estimated association was –3.2% (95% CI [–7.6, 1.2]). Among conservative individuals, the estimated association was –13.9% (95% CI [–33.8, 6.1]), while for liberals the effect was negligible (–0.3%; 95% CI [–2.2, 1.7]). For the dosage measure, the overall dosage association was 0.010% (95% CI [–0.020, 0.041]).

These findings provide additional evidence that the association between untrustworthy website exposure and false election beliefs is ideologically contingent. Exposure to conservative untrustworthy websites is associated with a significant positive association with false election beliefs, most notably among conservative individuals, whereas exposure to liberal untrustworthy websites appears to exert a null or even slightly negative association. This result aligns with the information environment of the 2020 US presidential election, where narratives questioning the election's legitimacy were a unique and central feature of conservative-leaning untrustworthy domains. The tight correspondence between the ideological source of the narrative and the statistical association provides evidence that the mechanism is exposure to the political content of these sites. These heterogeneous differences thus underscore the importance of considering the connection between content ideology and audience predispositions when evaluating untrustworthy content and its antecedents.

Table 3: Robustness Analysis: Alternative Ideology Thresholds for Conservative Domain Classification

Threshold	Model	Overall	Trump Supporters	Non-Trump Supporters
±0.1	Binary	0.0598 (0.0233)	0.1456 (0.0567)	0.0026 (0.0105)
		[0.0141, 0.1055]	[0.0346, 0.2567]	[-0.0180, 0.0232]
±0.2	Binary	0.0545 (0.0233)	0.1445 (0.0566)	-0.0064 (0.0080)
		[0.0089, 0.1001]	[0.0336, 0.2555]	[-0.0220, 0.0093]
±0.3	Binary	0.0508 (0.0232)	0.1357 (0.0563)	-0.0067 (0.0082)
		[0.0054, 0.0963]	[0.0253, 0.2460]	[-0.0228, 0.0094]
±0.4	Binary	0.0556 (0.0238)	0.1420 (0.0560)	-0.0063 (0.0085)
		[0.0089, 0.1023]	[0.0323, 0.2517]	[-0.0229, 0.0103]
±0.5	Binary	0.0631 (0.0274)	0.1296 (0.0552)	-0.0003 (0.0109)
		[0.0094, 0.1168]	[0.0214, 0.2378]	[-0.0217, 0.0210]
±0.6	Binary	0.0617 (0.0277)	0.1272 (0.0558)	-0.0006 (0.0111)
		[0.0075, 0.1159]	[0.0179, 0.2365]	[-0.0224, 0.0211]
±0.7	Binary	0.0720 (0.0312)	0.1369 (0.0582)	0.0002 (0.0139)
		[0.0109, 0.1332]	[0.0229, 0.2510]	[-0.0271, 0.0275]
±0.1	Dosage	0.0005 (0.0001)	0.0005 (0.0001)	0.0015 (0.0021)
		[0.0002, 0.0008]	[0.0002, 0.0008]	[-0.0027, 0.0057]
±0.2	Dosage	0.0005 (0.0001)	0.0005 (0.0001)	0.0015 (0.0023)
		[0.0003, 0.0008]	[0.0002, 0.0008]	[-0.0029, 0.0059]
±0.3	Dosage	0.0005 (0.0001)	0.0005 (0.0001)	0.0016 (0.0023)
		[0.0003, 0.0008]	[0.0002, 0.0008]	[-0.0029, 0.0061]
±0.4	Dosage	0.0005 (0.0001)	0.0005 (0.0001)	0.0017 (0.0024)
		[0.0003, 0.0008]	[0.0002, 0.0008]	[-0.0030, 0.0064]
±0.5	Dosage	0.0005 (0.0001)	0.0005 (0.0001)	0.0018 (0.0024)
		[0.0003, 0.0008]	[0.0002, 0.0008]	[-0.0030, 0.0066]
±0.6	Dosage	0.0005 (0.0001)	0.0005 (0.0001)	0.0020 (0.0025)
		[0.0003, 0.0008]	[0.0002, 0.0008]	[-0.0029, 0.0068]
±0.7	Dosage	0.0006 (0.0002)	0.0006 (0.0002)	0.0022 (0.0030)
		[0.0001, 0.0010]	[0.0001, 0.0010]	[-0.0038, 0.0081]

Note: This table presents robustness checks for alternative ideology threshold cutoffs used to classify conservative domains. The original analysis used thresholds of > 0 for conservative domains. Values represent associations with standard errors in parentheses and 95% confidence intervals in brackets. Results demonstrate that the main findings are robust across different threshold specifications.

Table 4: Robustness Analysis: Alternative Ideology Thresholds for Liberal Domain Classification

Threshold	**Model**	**Overall**	**Trump Supporters**	**Non-Trump Supporters**
±0.1	Binary	-0.0336 (0.0234)	-0.1608 (0.1162)	-0.0021 (0.0109)
		[-0.0794, 0.0122]	[-0.3885, 0.0669]	[-0.0233, 0.0192]
±0.2	Binary	-0.0375 (0.0234)	-0.1891 (0.1202)	-0.0021 (0.0109)
		[-0.0834, 0.0084]	[-0.4247, 0.0464]	[-0.0235, 0.0194]
±0.3	Binary	-0.0331 (0.0156)	-0.5979 (0.2319)	-0.0048 (0.0050)
		[-0.0637, -0.0025]	[-1.0525, -0.1434]	[-0.0146, 0.0050]
±0.4	Binary	-0.0331 (0.0156)	-0.5979 (0.2319)	-0.0048 (0.0050)
		[-0.0637, -0.0025]	[-1.0525, -0.1434]	[-0.0146, 0.0050]
±0.5	Binary	-0.0331 (0.0156)	-0.5979 (0.2319)	-0.0048 (0.0050)
		[-0.0637, -0.0025]	[-1.0525, -0.1434]	[-0.0146, 0.0050]
±0.6	Binary	-0.0331 (0.0156)	-0.5979 (0.2319)	-0.0048 (0.0050)
		[-0.0637, -0.0025]	[-1.0525, -0.1434]	[-0.0146, 0.0050]
±0.7	Binary	-0.0331 (0.0156)	-0.5979 (0.2319)	-0.0048 (0.0050)
		[-0.0637, -0.0025]	[-1.0525, -0.1434]	[-0.0146, 0.0050]
±0.1	Dosage	0.0001 (0.0002)	0.0009 (0.0037)	-0.0000 (0.0000)
		[-0.0003, 0.0006]	[-0.0064, 0.0083]	[-0.0000, 0.0000]
±0.2	Dosage	0.0001 (0.0002)	0.0009 (0.0030)	-0.0000 (0.0000)
		[-0.0003, 0.0006]	[-0.0049, 0.0067]	[-0.0000, 0.0000]
±0.3	Dosage	-0.0000 (0.0000)	-0.0140 (0.0275)	-0.0000 (0.0000)
		[-0.0001, 0.0000]	[-0.0679, 0.0400]	[-0.0000, 0.0000]
±0.4	Dosage	-0.0000 (0.0000)	-0.0140 (0.0275)	-0.0000 (0.0000)
		[-0.0001, 0.0000]	[-0.0679, 0.0400]	[-0.0000, 0.0000]
±0.5	Dosage	-0.0000 (0.0000)	-0.0140 (0.0275)	-0.0000 (0.0000)
		[-0.0001, 0.0000]	[-0.0679, 0.0400]	[-0.0000, 0.0000]
±0.6	Dosage	-0.0000 (0.0000)	-0.0140 (0.0275)	-0.0000 (0.0000)
		[-0.0001, 0.0000]	[-0.0679, 0.0400]	[-0.0000, 0.0000]
±0.7	Dosage	-0.0000 (0.0000)	-0.0140 (0.0275)	-0.0000 (0.0000)
		[-0.0001, 0.0000]	[-0.0679, 0.0400]	[-0.0000, 0.0000]

Note: This table presents robustness checks for alternative ideology threshold cutoffs used to classify liberal domains. The original analysis used thresholds of < 0 for liberal domains. Values represent associations with standard errors in parentheses and 95% confidence intervals in brackets. Results demonstrate that the main findings are robust across different threshold specifications.

Appendix F: Omitted Variable and Confounder Robustness Checks

We examine the robustness of the assumption of non-confoundedness via a causal sensitivity analysis. To operate under the conservative assumption that we do not know the specific parameters of a potential confounding variable (McGowan 2022), we calculate a Robustness Value (r-value) using the `tipr` R package (McGowan 2022). The Robustness Value (RV) represents the percentage of the residual variance that an unobserved confounder would have to explain of both the treatment and the outcome to "explain away" the observed difference. These RVs are reported in Table 5.

Table 5: Robustness values for significant causal forest results

Result	Robustness Value (RV)
Exposure	0.06
Exposure - Trump Supporters	0.11
Dosage	0.12
Dosage - Trump Supporters	0.17

Note: The Robustness Value (RV) represents the percentage of the residual variance that an unobserved confounder would need to explain (of both the treatment and the outcome) to "explain away" the observed difference.

To put these values into perspective, we go variable-by-variable in the dataset, removing each observed control variable in turn from the relationship estimation in the causal forest process. We then examine whether the removal of that variable would confound the analysis, given the RV for the analysis and the residual variance explained by the removed variable. The results are shown in Table 6.

Table 6: Confounder analysis

Variable	Exposure		Exposure - Trump Supporters		Dosage		Dosage - Trump Supporters	
	Res. Var.	Confounder	Res. Var.	Confounder	Res. Var.	Confounder	Res. Var.	Confounder
UntrustworthyWebsiteExposureWave1	0.040	No	0.001	No	0.000	No	0.006	No
Conservative	0.749	Yes	0.000	No	0.032	No	0.000	No
EduCollegeGrad	0.038	No	0.128	Yes	0.018	No	0.034	No
EduHSorLess	0.001	No	0.120	Yes	0.025	No	0.069	No
EduPostGrad	0.043	No	0.065	No	0.335	Yes	0.233	Yes
EduSomeCollege	0.114	Yes	0.525	Yes	0.428	Yes	0.425	Yes
Female	0.001	No	0.004	No	0.021	No	0.097	No
RaceBlack	0.028	No	0.001	No	0.004	No	0.000	No
RaceHispanic	0.004	No	0.009	No	0.004	No	0.009	No
RaceOther	0.000	No	0.005	No	0.001	No	0.004	No
RaceWhite	0.028	No	0.010	No	0.009	No	0.013	No
PoliticalKnowledge	0.000	No	0.019	No	0.127	Yes	0.157	No
PoliticalInterest	0.005	No	0.002	No	0.032	No	0.034	No
AgeUnder30	0.009	No	0.000	No	0.000	No	0.000	No
Age30to44	0.003	No	0.015	No	0.000	No	0.000	No
Age45to65	0.005	No	0.021	No	0.079	No	0.092	No
Age65plus	0.034	No	0.059	No	0.100	No	0.117	No

Note: Confounder analysis assessing whether a given observed control variable would confound the results if omitted from the analysis.

We test for omitted variable bias *under the correlation-based regression framework* using the method and robustness cutoff suggested by Oster (2019), as implemented by the `robomit` R package (Schaub and Zurich 2020). Specifically, we conduct bootstraps with replacement for 1,000 simulations with 1,000 draws each. The R_{max} value we use is the R^2 of the full controlled model multiplied by 1.3, as suggested by Oster (2019).

The bias-adjusted coefficients are shown in Table 7. Although this omitted variable bias robustness test is conducted on the regression analysis, the fact that the coefficients remain positive suggests that the results are robust to omitted variable bias. This test is noted as being a very conservative approach to testing for omitted variable bias, with only 45% of nonrandomized results in top economics journals passing this robustness test (Oster 2019).

Table 7: Omitted variable bias test

Model	$\hat{\beta}^{\text{base}}_{\text{ExposedUntrustworthy}}$	$\hat{\beta}^{\text{full}}_{\text{ExposedUntrustworthy}}$	$\hat{\beta}^{\text{bias-adjusted}}_{\text{ExposedUntrustworthy}}$
Binary Exposure	0.1733	0.0578	0.0001
Dosage Model	0.0008	0.0004	0.0002

Note: Omitted variable bias test as described in Oster (2019). The Model column distinguishes between the binary exposure and dosage models. The $\hat{\beta}^{\text{full}}_{\text{ExposedUntrustworthy}}$ column shows the coefficient from the full model in the Gelbach decompositions (see Table 1 and Table 2), and $\hat{\beta}^{\text{bias-adjusted}}_{\text{ExposedUntrustworthy}}$ is the bias-adjusted coefficient using Oster's method.

Furthermore, we also estimated the association between untrustworthy exposure and other political behaviors in Table 8. We do not find significant results for any of these behaviors, providing further evidence that our main analysis is robust to latent confounding.

Table 8: Other dependent variables

DV	Association	SE	CI	z-value	p-value
Support - Trump	0.019	0.015	[-0.011, 0.049]	1.240	0.215
Attend Rally - Trump	0.002	0.009	[-0.016, 0.019]	0.208	0.835
Attend Rally - Biden	-0.003	0.007	[-0.017, 0.011]	-0.447	0.655
Attend Rally - Other	0.016	0.010	[-0.003, 0.036]	1.621	0.105
Volunteer - Biden	-0.014	0.013	[-0.039, 0.010]	-1.128	0.259
Volunteer - Trump	-0.004	0.006	[-0.016, 0.008]	-0.626	0.531
Volunteer - Other	-0.007	0.011	[-0.029, 0.015]	-0.643	0.520
Volunteer - Poll Worker	-0.012	0.008	[-0.027, 0.004]	-1.436	0.151
Donation - Trump	-0.001	0.015	[-0.031, 0.029]	-0.088	0.930
Donation - Biden	-0.003	0.023	[-0.047, 0.042]	-0.123	0.902
Donation - Other	0.006	0.024	[-0.041, 0.053]	0.248	0.804
Social Media - Trump	0.028	0.020	[-0.012, 0.068]	1.387	0.165
Social Media - Biden	-0.014	0.026	[-0.065, 0.038]	-0.512	0.608
Social Media - Other	0.047	0.025	[-0.001, 0.095]	1.919	0.055
Yard Sign - Trump	0.010	0.015	[-0.019, 0.039]	0.654	0.513
Yard Sign - Biden	-0.023	0.017	[-0.057, 0.012]	-1.297	0.195
Yard Sign - Other	-0.018	0.017	[-0.052, 0.016]	-1.054	0.292

Note: Additional analysis of other outcome variables. DV = dependent variable; Association = average association; SE = standard error; CI = 95% confidence interval; z-value = z-value; p-value = p-value.