

A Signal Detection Analysis of Source Effects on Misinformation Susceptibility

Nyx L. Ng, Fabian M. Wurzinger, & Bertram Gawronski

University of Texas at Austin

in press, *Journal of Experimental Social Psychology*

Author Note

The authors thank Isabelle Azmoodeh, Bryanna Boone, Alexandra Castro, Ana Sofia Cerda Ibarrola, Audrey-Anna Scalici, and Alyssa Yarbrough for their help in preparing the stimulus materials. This research was supported by National Science Foundation Grant BCS-2040684. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors declare no conflict of interest. Correspondence concerning this article should be addressed to Nyx L. Ng (nyx.ng@nyu.edu) who is now at New York University.

Abstract

Signal Detection Theory suggests that people may fall for misinformation because they are unable to discern true from false information (i.e., low truth sensitivity), or because they have a general tendency to accept information regardless of veracity (i.e., low acceptance threshold). While a large body of evidence suggests that characteristics of information sources can influence susceptibility to misinformation, it remains unclear how source characteristics affect misinformation susceptibility. Using a signal-detection approach, the current research investigated (1) whether source characteristics influence truth judgments of true and false information via truth sensitivity or acceptance thresholds; (2) potential differences between perceived reliability and perceived partisanship as distinct facets of sources; and (3) how the two facets interact with information content in shaping judgments of true and false information. Across seven preregistered experiments ($N = 4,286$), we found that (1) source characteristics affected susceptibility to misinformation via acceptance thresholds, not truth sensitivity *per se*; (2) effects of source characteristics were limited to conditions where comparisons of sources are salient; (3) source characteristics had no effect on partisan bias, in that acceptance thresholds were substantially lower for ideology-congruent than ideology-incongruent information regardless of the source; and (4) perceived reliability and perceived partisanship showed corresponding effects on acceptance thresholds regardless of information content, suggesting that both source characteristics might influence truth judgments via perceived source credibility. Together, the findings provide deeper insights into how source characteristics influence susceptibility to misinformation.

Keywords: misinformation, partisanship, political beliefs, source characteristics, truth judgments

A Signal Detection Analysis of Source Effects on Misinformation Susceptibility

From examining factors underlying responses to true and false information (e.g., Gawronski et al., 2023; Pennycook & Rand, 2021; Pfänder & Altay, 2025; Schwarz & Jalbert, 2020; Sultan et al., 2024; Van Bavel et al., 2024) to testing interventions aimed at curbing the ballooning misinformation epidemic (e.g., Chan & Albarracín, 2023; Roozenbeek & van der Linden, 2019), psychological research on how people navigate misinformation has proliferated in recent years (for reviews, see Ecker et al., 2022; Pennycook & Rand, 2021). An important question in this work is how characteristics of information sources affect susceptibility to misinformation (i.e., judgments of false information as true). Prior research suggests that people are less likely to fall for misinformation from non-credible sources (e.g., Bauer & Clemm Von Hohenberg, 2021; Clemm Von Hohenberg & Guess, 2023; Nadarevic et al., 2020; Zeng et al., 2024). However, based on the available evidence, it remains unclear (1) how source characteristics influence truth judgments of true and false information, (2) whether distinct facets of sources (e.g., perceived reliability vs. perceived partisanship) influence judgments of true and false information in the same manner, and (3) how distinct facets of sources interact with information content in shaping judgments of true and false information. Answers to these questions are critical for understanding why people fall for misinformation, which has important implications for the effectiveness of interventions to reduce misinformation susceptibility.

In the current research, we drew on Signal Detection Theory (SDT; Green & Swets, 1966) to address these questions. According to SDT, people may fall for misinformation because they are unable to discern true from false information (i.e., low truth sensitivity; see Batailler et al., 2022), or because they have a general tendency to accept information regardless of veracity (i.e., low acceptance threshold; see Batailler et al., 2022). Expanding on this distinction, the

current research used a signal-detection approach to investigate whether source characteristics influence truth judgments of true and false information via truth sensitivity or acceptance thresholds. We further investigated whether perceived source reliability and perceived source partisanship influence susceptibility to misinformation in the same manner, and how the two facets of sources interact with information content in shaping judgments of true and false information.

Misinformation as a Signal Detection Problem

In signal-detection terms, correct judgments of true information as true can be described as *hits*; incorrect judgments of false information as true can be described as *false alarms*; correct judgments of false information as false can be described as *correct rejections*; and incorrect judgments of true information as false can be described as *misses* (Batailler et al., 2022). SDT provides a mathematical approach for quantifying two conceptually distinct factors underlying misinformation susceptibility.

First, truth judgments of true and false information are shaped by the ability to discern true from false information (i.e., *truth sensitivity*; see Batailler et al., 2022). High truth sensitivity reflects high ability at distinguishing true from false information, involving high rates of *hits* and *correct rejections* and, correspondingly, low rates of *misses* and *false alarms*. Conversely, low truth sensitivity reflects low ability at distinguishing true from false information, involving high rates of *misses* and *false alarms* and, correspondingly, low rates of *hits* and *correct rejections*. Truth sensitivity is captured by SDT's d' index, which reflects the distance between the distributions for true and false information along the judgment dimension of perceived veracity (see Figure 1). It is calculated as follows:

$$d' = z(H) - z(FA)$$

In this equation, H represents the proportion of true information judged as true; FA represents the proportion of false information judged as true. Both proportions are converted to follow a quantile function for a z -distribution, such that a proportion of 0.5 is converted to a d' score of zero (reflecting chance responses). Higher d' scores reflect greater accuracy in discerning true from false information (i.e., higher truth sensitivity).

Second, truth judgments of true and false information are shaped by general response tendencies to accept or reject information regardless of the information's veracity (i.e., *acceptance threshold*; see Batailler et al., 2022). In terms of SDT, this tendency can be understood as the threshold along the judgment dimension of perceived veracity at which a person switches their judgment between true and false (see Figure 2). A high acceptance threshold reflects a general tendency to reject information as false, involving high rates of *misses* and *correct rejections* and, correspondingly, low rates of *hits* and *false alarms*. Conversely, a low acceptance threshold reflects a general tendency to accept information as true, involving high rates of *hits* and *false alarms* and, correspondingly, low rates of *misses* and *correct rejections*. Acceptance threshold is captured by SDT's c index and is calculated as follows:

$$c = -0.5 \times [z(H) + z(FA)]$$

Acceptance threshold scores higher than zero indicate a greater likelihood to judge information as false than to judge information as true; scores lower than zero indicate a greater likelihood to judge information as true than to judge information as false. A score of zero reflects an equal likelihood to judge information as true versus false.

An important aspect of the distinction between truth sensitivity and acceptance threshold is that the two factors are independent (Macmillan & Creelman, 2004). For example, a high acceptance threshold does not entail high truth sensitivity, given that high thresholds imply not

only (1) a high likelihood of judging false information as false (i.e., high rate of correct rejections) but also (2) a high likelihood of judging true information as false (i.e., high rate of misses). In a similar vein, a low acceptance threshold does not entail low truth sensitivity, given that low thresholds imply not only (1) a high likelihood of judging false information as true (i.e., high rate of false alarms) but also (2) a high likelihood of judging true information as true (i.e., high rate of hits). While some factors may shape misinformation belief via both truth sensitivity and acceptance thresholds (e.g., older age; Sultan et al., 2024), not all exert concurrent and similar effects on both. For example, while cognitive elaboration has been found to affect truth sensitivity without affecting acceptance thresholds (Gawronski et al., 2023; Nahon et al., 2024; Sultan et al., 2022), effects of familiarity on misinformation belief (Pillai & Fazio, 2021) seem to stem from effects on acceptance thresholds and not truth sensitivity (Gawronski et al., 2026; Sultan et al., 2024). From an intervention standpoint, there are ongoing debates about whether gamified inoculation increases acceptance thresholds without affecting truth sensitivity (Modirrousta-Galian & Higham, 2023) or, conversely, increases truth sensitivity without affecting acceptance thresholds (Simchon et al., 2026). Thus, when studying why people mistakenly judge false information as true, it is critical to delineate the distinct roles of truth sensitivity and acceptance threshold (see Batailler et al., 2022; Gawronski et al., 2023; Hubeny et al., 2025; Modirrousta-Galian & Higham, 2023; Nahon et al., 2024; Simchon et al., 2026; Sultan et al., 2022).

Effects of Source Characteristics on Misinformation Susceptibility

Judgments of whether information is true or false are likely shaped not only by the content of the information but also by its source. Research on attitudes suggests a key role of source characteristics in persuasion, with people being more persuaded by information provided

by sources perceived to be credible (Chaiken, 1980; Petty et al., 1997; Petty & Cacioppo, 1986). While some studies found no or inconsistent effects of source credibility on misinformation susceptibility (Dias et al., 2020; Littrell et al., 2024; Pehlivanoglu et al., 2021; Schaewitz et al., 2020; see Mang et al., 2024 for a discussion on the factors potentially driving the inconsistent findings), several studies suggest that people are more likely to judge information as true when it is coming from a credible (vs. non-credible) source (e.g., Bauer & Clemm Von Hohenberg, 2021; Clemm Von Hohenberg & Guess, 2023; Nadarevic et al., 2020; Zeng et al., 2024). Further, studies on responses to contradictory information suggest that people are more likely to update their beliefs about an earlier allegation when it is contradicted by a trustworthy (vs. untrustworthy) source (e.g., Ecker & Antonio, 2021; Sanna & Lagnado, 2025).

From a practical view, these findings seem encouraging, because they suggest that people are less likely to believe information from non-credible sources. Because misinformation is often spread by non-credible sources (e.g., dubious or fake-news outlets; laypeople lacking expertise on an issue), this tendency may protect people from falling for misinformation (Schwarz & Jalbert, 2020). Extant research on misinformation susceptibility has often adopted procedures that reflect this idea, with studies presenting true information with credible sources and false information with non-credible sources (e.g., Aslett et al., 2024; Dias et al., 2020; Pennycook et al., 2021; Pennycook & Rand, 2019).

From an applied perspective, the confound of source credibility and information veracity in these studies may seem insignificant because (1) the truth status of information in natural settings is typically positively correlated with source credibility (i.e., information from credible sources is more likely to be true than information from non-credible sources) and, therefore, (2) the two factors should lead to the same desirable outcome: a lower likelihood of accepting

misinformation conveyed by non-credible sources (e.g., Swire-Thompson et al., 2024). However, from a theoretical perspective, study designs that confound source credibility and information veracity can lead to inaccurate conclusions about why people fall for misinformation. To illustrate this point, consider that false information can be spread not only by non-credible sources but also by credible sources (e.g., Peters, 2023), and that even non-credible sources may convey true information. If people readily accept information from credible sources and reject information from non-credible sources, they may mistakenly accept false information as true when it is conveyed by a credible source, and they may mistakenly reject true information as false when it is conveyed by a non-credible source. Thus, if one aims to understand why people fall for misinformation, it is critical to avoid confounds between source credibility and information veracity, because such confounds can lead to inaccurate theoretical conclusions.

Some studies have isolated effects of source characteristics and information veracity on truth judgments (Nedelcu & Balaban, 2021; Pehlivanoglu et al., 2021; Traberg et al., 2024; Traberg & van der Linden, 2022). However, because these studies did not delineate effects on truth sensitivity and acceptance thresholds, it remains unclear how source characteristics affected truth judgments. While past findings that people are more likely to accept information from credible than non-credible sources may suggest that source characteristics influence general tendencies to accept or reject information (i.e., acceptance thresholds; see Traberg & van der Linden, 2022), source characteristics may also influence discernment between true and false information (i.e., truth sensitivity; see Sultan et al., 2024). Effects of the latter type may be expected based on research showing that sources with low credibility can increase elaborate processing (Priester & Petty, 1995). Combined with evidence that elaborate processing increases accuracy in discerning true from false information (for a review, see Pennycook & Rand, 2021),

these findings suggest that people may show higher truth sensitivity for information from non-credible sources compared to information from credible sources.

Notably, a meta-analysis by Sultan and colleagues (2024) has used an SDT approach to investigate whether source characteristics affect misinformation susceptibility via truth sensitivity or acceptance thresholds. The authors found that truth sensitivity was higher when true and false headlines were presented with their sources than when they were presented without sources. Presence (vs. absence) of sources had no effect on acceptance thresholds. While these findings seem to suggest that source characteristics influence misinformation susceptibility via truth sensitivity and not via acceptance thresholds, most of the studies in the meta-analysis that included sources confounded source characteristics and information veracity, in that true information was presented with credible sources whereas false information was presented with non-credible sources (e.g., Pennycook et al., 2021). This confound renders conclusions about source effects on truth sensitivity premature.

In terms of SDT, presenting true information with a credible source should increase the proportion of hits, whereas presenting false information with a non-credible source should decrease the proportion of false alarms (see Table 1, upper section). These effects should lead to higher truth-sensitivity scores when source information is present (vs. absent). However, the confound between information veracity and source credibility obfuscates conclusions about whether source credibility affects truth sensitivity *per se*. To address this question, it is necessary to resolve the confound between information veracity and source reliability in a fully crossed 2 (Accuracy: true vs. false) $\times$ 2 (Source: credible vs. non-credible) design that includes (1) true information paired with a credible source, (2) false information paired with a non-credible source, (3) true information paired with a non-credible source, and (4) false information paired

with a credible source (see Table 1, lower section). In such a design, one would again expect higher hit rates for true information paired with a credible source and lower false-alarm rates for false information paired with a non-credible source. Yet, if people readily accept information from credible sources, they should also exhibit higher false-alarm rates for information from credible sources, such that they mistakenly judge false information from credible sources to be true. By the same token, if people readily reject information from non-credible sources, they should exhibit lower hit rates for information from non-credible sources, such that they mistakenly judge true information from non-credible sources to be false. In conjunction, these effects imply that source credibility may not affect truth sensitivity *per se*, but instead influence acceptance thresholds, such that people show lower acceptance thresholds for information from credible sources compared to information from non-credible sources. Thus, despite the use of SDT in Sultan et al.'s (2024) meta-analysis, the confound between source information and information veracity in the meta-analyzed studies undermines conclusions about whether source credibility affects misinformation susceptibility via truth sensitivity or acceptance thresholds. The current research sought to address this question.

Perceived Reliability vs. Perceived Partisanship

Prior research examining effects of source characteristics on misinformation susceptibility has operationalized source credibility in terms of whether the source (1) is an expert or a lay person (Nadarevic et al., 2020; Sanna & Lagnado, 2025), (2) is real or fake (Bauer & Clemm Von Hohenberg, 2021; Nadarevic et al., 2020), or (3) is a mainstream or fringe source (Nedelcu & Balaban, 2021). A related source characteristic is a source's political leaning (Traberg et al., 2024), which can increase or decrease perceived source credibility depending on whether a source's political leaning is congruent or incongruent with participants' own political

leaning (Traberg & van der Linden, 2022). Because different source characteristics may differentially impact judgments (e.g., Ecker & Antonio, 2021; Ziegler et al., 2002), comparing effects of different source characteristics is crucial for understanding how source information affects susceptibility to misinformation (Mang et al., 2024). In the current research, we tested potential differences between perceived reliability and perceived partisanship. We focused on these two source characteristics, because we expected them to show distinct effects on acceptance thresholds depending on the content of the focal information.

Perceived source reliability can be understood as the subjective likelihood that a given source conveys true (vs. false) information. Information from a source with high perceived reliability is likely trusted more than information from a source with low perceived reliability, and this may be the case irrespective of the information content, such as whether the political slant of the information aligns with one's political views (henceforth referred to as ideology-congruent vs. ideology-incongruent information). In other words, people may show lower acceptance thresholds for information from reliable sources than for information from unreliable sources, regardless of whether the information is ideology-congruent, ideology-incongruent, or apolitical.

In contrast, perceived source partisanship can be understood as the subjective likelihood that a given source conveys information that favors one political side over another (e.g., a source that is perceived to favor pro-Republican over pro-Democrat information, or vice versa). Different from the content-independent effects of perceived reliability, effects of perceived partisanship likely depend on whether the political slant of the information aligns with the recipient's political views. On the one hand, the low subjective likelihood of ideology-congruent information being reported by an ideology-incongruent source may lead recipients to assume that

the information must be true—otherwise the source would not report it. Thus, while people tend to show lower acceptance thresholds for ideology-congruent than ideology-incongruent information (Batailler et al., 2022; Gawronski et al., 2023), their threshold for accepting ideology-congruent information may be even lower when the information comes from an ideology-incongruent source (e.g., a Democrat seeing pro-Democrat information reported on Fox News; a Republican seeing pro-Republican information reported on CNN) than when it comes from an ideology-congruent source (e.g., a Democrat seeing pro-Democrat information reported on CNN; a Republican seeing pro-Republican reported on Fox News). In terms of classic attribution theory (Kelley & Michela, 1980), one could say that inconsistency between the political slant of information and the political leaning of its source augments the diagnostic value of the information, whereas consistency between the two suggests that its diagnostic value should be discounted.

On the other hand, the low subjective likelihood of ideology-incongruent information being reported by an ideology-congruent source may similarly lead recipients to assume that the information must be true, following the attributional principles of augmentation and discounting (Kelley & Michela, 1980). Thus, different from the hypothesized pattern for ideology-congruent information, people's threshold for accepting ideology-incongruent information may be lower when the information comes from an ideology-congruent source (e.g., a Democrat seeing pro-Republican information on CNN; a Republican seeing pro-Democrat information on Fox News) than when it comes from an ideology-incongruent source (e.g., a Democrat seeing pro-Republican information on Fox News; a Republican seeing pro-Democrat information on CNN).

Together, these considerations suggest that effects of perceived partisanship may depend on the information content, in that (1) acceptance thresholds for ideology-congruent information

should be lower for ideology-incongruent than ideology-congruent sources, and conversely (2) acceptance thresholds for ideology-incongruent information should be lower for ideology-congruent than ideology-incongruent sources. Moreover, because perceived partisanship involves expectancies that are specific to political information, it may have no effect for apolitical information. These predictions are different from the ones for perceived reliability, which can be expected to influence acceptance thresholds irrespective of the information content. The current research sought to test these predictions.

The Current Research

The three objectives of the current research were to investigate (1) whether source characteristics influence truth judgments of true and false information via truth sensitivity or acceptance thresholds, (2) potential differences between perceived reliability and perceived partisanship as distinct facets of sources, and (3) how each facet interacts with information content in shaping judgments of true and false information. To this end, we conducted seven preregistered experiments in which self-identified Democrats and self-identified Republicans judged the truth of true and false news headlines with either a pro-Democrat or a pro-Republican slant. In Experiments 1-4, the headlines were presented with a reliable or an unreliable source (i.e., Forbes vs. National Enquirer). In Experiments 5-6, the headlines were presented with a Democrat-leaning or Republican-leaning source (i.e., CNN vs. Fox News). In all studies, we used SDT to delineate source effects on truth sensitivity and acceptance thresholds.

Experiment 1 used a confounded manipulation that mirrored the comparison in Sultan et al.'s (2024) meta-analysis, in that for half of the participants true headlines were always paired with a reliable source and false headlines were always paired with an unreliable source; the remaining half were presented with the same new headlines without source information. We

expected to replicate Sultan et al.'s meta-analytic finding that presence (vs. absence) of source information increases truth sensitivity when source reliability is confounded with information veracity.

Experiments 2a and 2b aimed to show that the effect on truth sensitivity in Experiment 1 is a spurious product of the confound between source reliability and information veracity (see Table 1). To this end, the two experiments used a fully crossed design in which all headlines were presented with either (1) a reliable source, (2) an unreliable source, or (3) no source. In this fully crossed design, we expected source reliability to influence acceptance thresholds (rather than truth sensitivity), in that acceptance thresholds should be lower when the headlines were presented with a reliable source than when they were presented with an unreliable source.

Expanding on unexpected outcomes of Experiments 2a and 2b, Experiment 3 used an extended design that included apolitical headlines in addition to politically charged headlines with a pro-Democrat or a pro-Republican slant. Experiment 4 used a within-subjects manipulation of source reliability instead of the between-subjects manipulation in the preceding experiments. Experiments 5 and 6 are conceptually identical to Experiments 3 and 4, the only difference being that these studies tested effects of perceived source partisanship. Prior to the seven experiments, we conducted a pilot study to identify suitable news sources for the manipulations of perceived source reliability and perceived source partisanship.

Open Research Practices

Except for the pilot study, all experiments were preregistered prior to data collection. Power analyses were conducted using GPower 3.1.9.7 (Faul et al., 2007). We report how we determined our sample size as well as all data exclusions, manipulations, and measures. All preregistrations (where applicable), data, analysis codes, and materials of all experiments

(including the pilot tests) are available at <https://osf.io/2mnrc/>. In addition to reporting the preregistered analyses, we also report Bayes factor analyses for the interpretation of null results. Bayes factor analyses were conducted using JASP version 0.19.1. All studies reported here received ethical approval from the Institutional Review Board of the University of Texas at Austin under protocol STUDY0000822.

Pilot Study

To identify suitable news sources that could be used for the experiments, we pilot-tested 45 sources that ranged from mainstream news sources (e.g., Forbes, AP News) to tabloids and fringe sources (e.g., National Enquirer, Real News Network). The sources were largely United-States-centric, except for some major international outlets (see Supplementary Materials, Table S1, for the complete list of sources). For Experiments 1 to 4, we had aimed to identify (1) one news source that both Democrats and Republicans are familiar with, that both groups perceive as non-partisan, and that both groups perceive to be reliable in conveying accurate information, and (2) another news source that both Democrats and Republicans are familiar with, that both groups perceive as non-partisan, and that both groups perceive to be unreliable in conveying accurate information. For Experiments 5 and 6, we had aimed to identify (1) one news source that both Democrats and Republicans are familiar with and perceive to be pro-Democrat, and (2) another news source that both Democrats and Republicans are familiar with and perceive to be pro-Republican. Table 2 presents information on participant demographics; Table 3 presents the means and standard deviations of the perceived reliability, partisanship, and familiarity scores of the four chosen sources. A detailed description of the methods and the results of the Pilot Study is provided in the Supplementary Materials. The pilot study's data were collected in September 2022.

Based on the results of the Pilot Study, we chose Forbes as the reliable non-partisan news source and National Enquirer as the unreliable non-partisan news source to be included in Experiments 1 to 4 (see Supplementary Materials for further details). Forbes is an American media source focused primarily on business topics, whereas National Enquirer is an American tabloid source. For Experiments 5 and 6, we chose CNN as the pro-Democrat source and Fox News as the pro-Republican source (see Supplementary Materials for further details). Both CNN and Fox News are cable news networks with major followings in the United States.

Notably, source perceptions were confirmed only through the pilot test and not again measured in the main experiments. We deemed it problematic to measure perceived source characteristics in the main experiments, because (1) doing so before the truth-judgment task would enhance the salience of the relevant source characteristics in the truth-judgment task (which could artificially inflate the to-be-tested effects), and (2) doing so after the truth-judgment task could lead to spill-over effects from the headlines in the truth-judgment task (i.e., equal numbers of true pro-Democrat, false pro-Democrat, true Pro-Republican, and false pro-Republican headlines) on participants' judgments of the relevant source characteristics (i.e., reliability, partisanship).

Experiment 1

Experiment 1 examined how a confound between source reliability and information veracity affects truth sensitivity and acceptance thresholds. Toward this end, we procedurally reproduced the confound often present in misinformation research, such that true headlines were always presented with a reliable non-partisan source (i.e., Forbes), whereas false headlines were always presented with an unreliable non-partisan source (i.e., National Enquirer). In a control condition, participants were presented with the same headlines without source information. We

hypothesized that source information containing the confound between source reliability (reliable vs. unreliable) and information veracity (true vs. false) would increase truth sensitivity in discerning true from false information (Hypothesis 1), replicating meta-analytic findings by Sultan et al. (2024).

Going beyond the null effect on overall acceptance thresholds in Sultan et al.'s (2024) meta-analysis, we also investigated source effects on acceptance thresholds for ideology-congruent and ideology-incongruent information. Doing so allowed us to investigate effects on partisan bias, defined as the tendency to show lower acceptance thresholds for ideology-congruent than ideology-incongruent information (see Batailler et al., 2022). Assuming that people are more receptive to information provided by a reliable non-partisan source than an unreliable non-partisan source, we expected the presence of source information to reduce partisan bias, such that the difference in acceptance thresholds for ideology-congruent and ideology-incongruent headlines would be smaller when source information (confounded with information veracity) is present than when it is absent (Hypothesis 2). Details of Experiment 1's preregistered approach can be found at <https://osf.io/ks4p3/>. Experiment 1's data were collected in November 2022.

Method

Design. Experiment 1 employed a 2 (Headline Accuracy: true vs. false) $\times$ 2 (Headline Slant: pro-Democrat vs. pro-Republican) $\times$ 2 (Political Affiliation: Democrat vs. Republican) $\times$ 2 (Source: absent vs. present) mixed design, with the first two factors varying within-subjects and the latter two factors varying between-subjects. Participants were randomly assigned to one of two experimental conditions. In the source-absent condition, participants judged the veracity of political headlines that were presented without source information. In the source-present

condition, true headlines were presented with the Forbes logo (i.e., reliable source) whereas false headlines were presented with the National Enquirer logo (i.e., unreliable source).

Participants. Participants were recruited on Prolific Academic. To ensure comparable representation of Democrats and Republicans, we used Prolific's prescreening information to create two separate assignments: one that restricted participation to Prolific workers who identified as Democrats and another that restricted participation to Prolific workers who identified as Republicans. Additional prescreening filters were used to restrict participation to Prolific workers who (1) are 18 years or older, (2) have an approval rating of over 95% on prior assignments on Prolific, (3) are a citizen of the United States, (4) are a resident of the United States, (5) have completed at least 100 prior assignments on Prolific, (6) are fluent in English, and (7) have not participated in prior studies from our lab using the same materials.

We aimed to have at least 80% power for the detection of a small between-group effect of $f = 0.10$ in a 2 (Source Information: absent vs. present) $\times$ 2 (Headline Slant: ideology-congruent vs. ideology-incongruent) mixed ANOVA with the first variable as a between-subjects factor and the second one as a within-subjects factor (two-tailed). Assuming a correlation between measures of $r = .30$ and using a nonsphericity correction of $\epsilon = 1$, a priori power analysis suggested a sample size of at least 278 participants is required to meet this goal. Anticipating that approximately 15% of the participants would be excluded based on preregistered exclusion criteria and to obtain sufficient power for slightly smaller effects, we aimed for a sample of 400 participants prior to preregistered exclusions. Data collection thus ended once 400 participants (200 per political affiliation) were approved to receive the US-\$3.00 compensation.

We received 203 responses from self-identified Republicans and 202 from self-identified Democrats, but only 200 per sample completed all measures and were thus compensated. Of

these, 195 Republicans and 199 Democrats passed the attention check, but one participant in the Republican sample reported inconsistent political affiliations in Prolific's prescreening survey and the study's item on political affiliation. The final sample is thus $N = 393$. Table 2 provides information on participant demographics.

Measures. The 60 headlines included in the study varied in terms of whether they (1) are true or false and (2) have a pro-Democrat or pro-Republican slant, with each of the four categories having 15 headlines. The headlines were selected based on fact-checks and pilot-tests for perceived familiarity and partisanship by Gawronski et al. (2023). The pilot-tests were conducted using balanced Prolific samples of Republicans and Democrats. Fifty-five headlines included in the current research were taken from Gawronski et al.'s (2023) second pilot test and five were taken from their first pilot test. Not all headlines were taken from the same pilot tests because some headlines within each set seemed outdated by the time of the current research. Additional details about the procedure for headline selection can be found in the Supplementary Materials. To standardize the presentation of the headlines, all headlines were presented in white font against a black background. Source information, where applicable, was depicted above each headline with the logo of either Forbes or the National Enquirer. For each headline, participants answered the question "*To the best of your knowledge, is the claim in this headline true or false?*" on a binary *True/False* scale.

Procedure. Participants first read the consent form and responded to two submission policy items. Next, they completed the headline judgment task. Participants in the source-present condition saw 30 true headlines paired with the reliable source *Forbes* and 30 false headlines paired with the unreliable source *National Enquirer*. Participants judged whether the claim in each headline is true or false (one at a time), with the headline order randomized per participant.

After judging all 60 headlines, participants answered questions assessing their political ideology, political affiliation, basic demographics, and finally, a reading-intensive attention check that required them to not select any response options. The attention check was as follows:

To facilitate our research on decision-making we are interested in learning a little more about you, the decision-maker. Psychological research using text-based materials requires that study participants read the materials and do not skip over longer pieces of text. We are therefore interested in whether you actually take the time to read the directions; if not, then some of our manipulations that rely on changes in the instructions will be ineffective. To demonstrate that you have read the instructions, please ignore the question and response options below. Instead, simply continue to the next page without answering the question. Thank you very much.

Of the following destinations, which one would be your first choice for a vacation if you had a free all-inclusive round trip after the Covid-19 pandemic? (Check all that apply)'.

The study took approximately 15 minutes to complete.

Data Analysis

Data Exclusions. We preregistered that we would exclude participants from the final sample if they (1) did not complete the entire study, (2) failed the attention check, or (3) provided information on our measure of political affiliation that is inconsistent with their self-reported political affiliation in Prolific's prescreening survey.

Indices. Hit rates were calculated as the proportion of true headlines judged as true, with one hit rate calculated for headlines that are congruent with participants' political affiliation and another hit rate calculated for headlines incongruent with participants' political affiliation. False-alarm rates were calculated as the proportion of false information judged as true, with false-

alarm rates calculated separately for ideology-congruent and ideology-incongruent headlines as well. We preregistered that, when the proportion of “true” responses within a headline-category was either 0 or 1, we would follow MacMillan and Creelman’s (2004) recommendation of converting 0 to $1/2N$, and 1 to $1 - 1/2N$, with N being the number of trials per headline category. Indices of truth sensitivity (d') and acceptance threshold (c) were calculated separately for ideology-congruent and ideology-incongruent headlines using the respective hit and false-alarm rates.

Results

Truth Sensitivity. The preregistered 2 (Source Information: absent vs. present) $\times 2$ (Headline Slant: ideology-congruent vs. ideology-incongruent) mixed ANOVA with truth-sensitivity scores as the criterion yielded a significant main effect of Source Information, $F(1, 391) = 8.56, p = .005, \eta_p^2 = .021, BF_{10} = 2.72$, indicating that participants in the source-present condition showed significantly higher truth sensitivity than those in the source-absent condition (see Table 4). The main effect of Headline Slant and its interaction with Source Information were not significant, $F_s \leq 2.04, p_s > .154, \eta_p^2_s < .005, BF_{10s} < 0.15$.

Acceptance Threshold. The preregistered 2 (Source Information: absent vs. present) $\times 2$ (Headline Slant: ideology-congruent vs. ideology-incongruent) mixed ANOVA with acceptance-threshold scores as the criterion (see Figure 3) yielded a significant main effect of Headline Slant, $F(1, 391) = 395.30, p < .001, \eta_p^2 = .503, BF_{10} = 4.88 \times 10^{13}$, indicating that participants showed significantly lower acceptance thresholds for ideology-congruent headlines than ideology-incongruent headlines. The main effect of Source Information and its interaction with Headline Slant were not significant, $F_s \leq 0.79, p_s > .380, \eta_p^2_s < .002, BF_{10s} < 0.08$.

Discussion

Consistent with Hypothesis 1, providing participants with source information that confounds source reliability with information veracity significantly increased truth sensitivity, in that participants showed higher accuracy in discerning true from false information when true information was presented with a reliable source and false information was presented with an unreliable source. Contrary to Hypothesis 2, source information did not reduce partisan bias, in that participants were significantly more likely to accept ideology-congruent information than ideology-incongruent information regardless of whether they judged the headlines with or without source information. Together, these results suggest that, when source reliability is confounded with information veracity, source reliability shapes truth judgments via truth sensitivity.

The obtained effect of source information on truth sensitivity is consistent with meta-analytic findings of Sultan et al. (2024), who found higher truth-sensitivity scores in studies that presented true and false headlines with their original sources compared to studies that presented true and false headlines without source information. Yet, as explained earlier, these findings do not necessarily imply that source information increases participants' ability to discern true from false information. After all, it is possible that participants merely have differential thresholds for information from credible and non-credible sources, which would produce a spurious effect on truth sensitivity when true information is presented with a credible source and false information is presented with a non-credible source. The main goal of the following experiments was to address this ambiguity.

Experiment 2a and 2b

In Experiments 2a and 2b, we aimed to disambiguate the effect of source reliability on truth judgments. Toward this end, we eliminated the confound between source reliability and information veracity by presenting all true and false headlines with either (1) a reliable source, (2) an unreliable source, or (3) no source, using a fully crossed design. We hypothesized that, when the confound is eliminated, truth sensitivity would not significantly differ across source conditions where information is presented with either a reliable source, an unreliable source, or no source (Hypothesis 1). Instead, we expected acceptance thresholds to significantly differ across source conditions (Hypothesis 2a), such that thresholds would be significantly higher in the unreliable-source condition compared to the reliable-source condition (Hypothesis 2b). In addition, we hypothesized that thresholds would be significantly lower in the no-source condition compared to the unreliable-source condition (Hypothesis 2c) and significantly higher compared to the reliable-source condition (Hypothesis 2d). Based on the results of Experiment 1, we further predicted that acceptance thresholds would be significantly lower for ideology-congruent than ideology-incongruent headlines (Hypothesis 3a), and that this partisan-bias effect would not significantly differ across source conditions (Hypothesis 3b). Because Experiment 2a yielded some unexpected results that are relevant for our hypotheses about effects of source reliability, we conducted a higher-powered, direct replication (Experiment 2b) to gauge the reliability of the obtained results. We present the results of both studies concurrently here. The design, hypotheses, and analysis plan of Experiments 2a and 2b were preregistered prior to data collection at <https://osf.io/we25a/> and <https://osf.io/vk2zf/>, respectively. Experiment 2a and 2b's data were collected in December 2022 and January 2023, respectively.

Method

Design. Experiments 2a and 2b employed a 2 (Headline Accuracy: true vs. false) $\times$ 2 (Headline Slant: pro-Democrat vs. pro-Republican) $\times$ 2 (Political Affiliation: Democrat vs. Republican) $\times$ 3 (Source Reliability: reliable vs. unreliable vs. no source) mixed design, with the first two factors varying within-subjects and the latter two factors varying between-subjects. Participants were randomly assigned to judge all true and false headlines presented with either a reliable source, an unreliable source, or no source.

Participants. We recruited participants using the same procedures as Experiment 1. Our aim was to have at least 80% power for the detection of a small between-group effect of $f = 0.10$ in a 3 (Source Reliability: reliable vs. unreliable vs. no source) $\times$ 2 (Headline Slant: ideology-congruent vs. ideology-incongruent) mixed ANOVA with the first factor varying between-subjects and the second varying within-subjects (two-tailed). Assuming a correlation between measures of $r = .30$ and using non-sphericity correction of $\epsilon = 1$, a sample size of at least 342 participants is required to meet this goal. Anticipating that approximately 15% of the participants would be excluded based on preregistered exclusion criteria and to obtain sufficient power to detect slightly smaller effects, we aimed for a sample of 600 participants prior to exclusions (300 per political affiliation) in Experiment 2a. Because we did not find the predicted source effects in Experiment 2a, we ran a follow-up experiment with the same design but with a larger sample to rule out the possibility that there was insufficient statistical power to detect source effects in Experiment 2a. To this end, we aimed for a sample of 900 participants prior to exclusions in Experiment 2b (450 per political affiliation). Anticipating that approximately 15% of the participants would be excluded based on preregistered exclusion criteria, a sample of 765 after exclusions provides 80% power for the detection of a small between-group effect of $f = 0.091$ in

a 3 (Source Reliability: reliable vs. unreliable vs. no source) $\times$ 2 (Headline Slant: ideology-congruent vs. ideology-incongruent) mixed ANOVA with the first variable as a between-subjects factor and the last one as a within-subjects factor, assuming a correlation between measures of $r = .30$ and using non-sphericity correction of $\epsilon = 1$. In both experiments, data collection ended once we approved compensation for the preregistered number of participants. Participants in both experiments received US-\$3.00 for their participation.

Experiment 2a. Six-hundred-and-six participants started the study. Two failed to pass a bot-check and another four did not complete the entire study. These participants were thus not compensated. Of the remaining 600 participants, 26 participants failed the attention check and another four provided inconsistent information about their political affiliation. The final sample was thus $N = 570$. Table 2 provides information on participant demographics.

Experiment 2b. Nine-hundred-and-fourteen participants started the study. Four failed to pass a bot-check and another nine did not complete the entire study, leaving us with 901 participants prior to exclusions, 900 of whom requested compensation. Fifty-six participants failed the attention check and another 10 provided inconsistent information about their political affiliation. The final sample was thus $N = 835$. Table 2 provides information on participant demographics.

Measures. Participants judged the same 60 pre-tested headlines as Experiment 1.

Procedure. The procedures of both experiments were largely identical to Experiment 1, with the exception that participants were randomly assigned to three conditions, where all headlines were presented either with (1) a reliable source (i.e., *Forbes*), (2) an unreliable source (i.e., *National Enquirer*), or (3) no source. The order of the 60 headlines was randomized individually for each participant. Each study took approximately 15 minutes to complete.

Data Analysis. Data exclusions and analytic approaches followed that of Experiment 1.

Results

Truth Sensitivity. The preregistered 3 (Source Reliability: reliable vs. unreliable vs. no source) $\times$ 2 (Headline Slant: ideology-congruent vs. ideology-incongruent) mixed ANOVA with truth sensitivity as the criterion yielded no significant main effect of Source Reliability on truth sensitivity in Experiment 2a, $F(1, 567) = 2.61, p = .075, \eta_p^2 = .009, BF_{10} = 0.14$, and Experiment 2b, $F(1, 832) = 0.32, p = .730, \eta_p^2 = .001, BF_{10} = 0.01$. A significant main effect of Headline Slant in Experiment 2a, $F(1, 567) = 5.61, p = .018, \eta_p^2 = .010, BF_{10} = 0.68$, and Experiment 2b, $F(1, 832) = 4.65, p = .031, \eta_p^2 = .006, BF_{10} = 0.36$, indicated that participants were better at discerning true from false information when the information was ideology-congruent than when it was ideology-incongruent (see Table 4). The main effect of Headline Slant was not moderated by Source Reliability in both experiments, $F_s \leq 0.44, p_s > .64, \eta_p^2_s < .002, BF_{10s} < 0.005$.

Acceptance Threshold. The preregistered 3 (Source Reliability: reliable vs. unreliable vs. no source) $\times$ 2 (Headline Slant: ideology-congruent vs. ideology-incongruent) mixed ANOVA with acceptance thresholds as the criterion (see Figure 4) yielded a significant main effect of Headline Slant in both Experiment 2a, $F(1, 567) = 602.90, p < .001, \eta_p^2 = .515, BF_{10} = 2.61 \times 10^{14}$, and Experiment 2b, $F(1, 832) = 780.53, p < .001, \eta_p^2 = .484, BF_{10} = 1.07 \times 10^{14}$. In both experiments, acceptance thresholds were significantly lower for ideology-congruent headlines than ideology-incongruent headlines. The main effect of Source Reliability was not statistically significant in both Experiment 2a, $F(1, 567) = 1.40, p = .247, \eta_p^2 = .005, BF_{10} = 0.04$, and Experiment 2b, $F(1, 832) = 2.67, p = .070, \eta_p^2 = .006, BF_{10} = 0.09$. The interaction between Source Reliability and Headline Slant was also not statistically significant in both Experiment 2a,

$F(2, 567) = .12, p = .884, \eta_p^2 < .001, BF_{10} = 0.00$, and Experiment 2b, $F(2, 832) = .26, p = .769, \eta_p^2 < .001, BF_{10} = 0.01$.

Discussion

Experiments 2a and 2b suggest that truth sensitivity does not significantly differ when political information is presented with a reliable source, an unreliable source, or no source, thus supporting Hypothesis 1. In other words, the effect of source information on truth sensitivity in Experiment 1 could have surfaced because of the incomplete experimental design that confounded source reliability and information veracity (see Table 1). However, counter to Hypothesis 2, acceptance thresholds also did not significantly differ across the three source conditions. Nevertheless, a significant partisan-bias effect emerged, in that participants showed lower acceptance thresholds for ideology-congruent information than for ideology-incongruent information regardless of source information, supporting Hypothesis 3.

Experiment 3

In trying to make sense of the unexpected null effects of source reliability in Experiments 2a and 2b, we wondered whether the extremely large partisan-bias effects nullified the predicted effect of source reliability on acceptance thresholds. If that is the case, source reliability should produce the expected effects on acceptance thresholds for apolitical information, but not for politically charged information. To test this possibility, Experiment 3 included apolitical headlines in the headline judgment task. Based on the findings of the previous experiments, we hypothesized that acceptance thresholds would significantly differ across different types of headlines (Hypothesis 1a), with acceptance thresholds being significantly lower for ideology-congruent than ideology-incongruent information (Hypothesis 1b). We further predicted that effects of source reliability on acceptance thresholds would depend on the headlines' political

slant (Hypothesis 2a), such that acceptance thresholds will be significantly lower for reliable than unreliable sources when the headlines are apolitical (Hypothesis 2b), but not when the headlines are ideology-congruent (Hypothesis 2c) or ideology-incongruent (Hypothesis 2d). Because the primary objective of Experiment 3 was to further examine the effect of source reliability on acceptance thresholds, we did not preregister any hypotheses related to truth sensitivity. That being said, for the sake of comprehensiveness, we still conducted preregistered exploratory analyses of potential effects on truth sensitivity. The design, hypotheses, and analysis plan of Experiment 3 were preregistered prior to data collection at <https://osf.io/9esbr/>. Experiment 3's data were collected in August 2023.

Method

Design. Experiment 3 employed a 2 (Headline Accuracy: true vs. false) $\times$ 3 (Headline Slant: pro-Democrat vs. pro-Republican vs. apolitical) $\times$ 2 (Political Affiliation: Democrat vs. Republican) $\times$ 3 (Source Reliability: reliable vs. unreliable vs. no source) mixed design, with the first two factors varying within-subjects and the latter two factors varying between-subjects. Like Experiments 2a and 2b, participants were randomly assigned to judge all headlines paired with either (1) a reliable source (i.e., Forbes), (2) an unreliable source (i.e., National Enquirer), or (3) no source.

Participants. We aimed for a sample of 900 participants prior to exclusions (450 per political affiliation). Anticipating that approximately 15% of the participants would be excluded based on preregistered exclusion criteria, a sample of 765 after exclusions provides 80% power for the detection of a small interaction effect of $f = 0.060$ in a 3 (Source Reliability: reliable vs. unreliable vs. no source) $\times$ 3 (Headline Slant: ideology-congruent vs. ideology-incongruent vs. apolitical) mixed ANOVA with the first variable as a between-subjects factor and the last one as

a within-subjects factor, assuming a correlation between measures of $r = .30$ and using non-sphericity correction of $\epsilon = 1$. Participants who completed the study were compensated US-\$3.00.

Nine-hundred-and-ten participants started the study. Two failed to pass a bot-check and another eight did not complete the entire study. The final sample we compensated thus comprised 900 participants. Eighteen failed the attention check and another six provided inconsistent information on their political affiliation. The final sample was thus $N = 876$. Table 2 provides information on participant demographics.

Measures. The 60 headlines included in the study varied in terms of whether they (1) are true or false and (2) are pro-Democrat, pro-Republican, or apolitical, with each of the six categories having 10 headlines. To obtain a set of up-to-date political news headlines and a set of apolitical headlines, we pilot-tested 179 headlines following the procedures of Gawronski et al. (2023). Details of the pilot-test and the selection criteria applied to identify headlines for inclusion as stimuli can be found in the Supplementary Materials.

Procedure. The procedures of Experiment 3 followed the ones in Experiments 2a and 2b with the exception that participants also judged apolitical headlines. The order of the 60 headlines was randomized individually for each participant. The study took approximately 15 minutes to complete.

Data Analysis. Data exclusions and analytic approaches followed those of the previous experiments, with the exception that d' and c scores were separately calculated for ideology-congruent, ideology-incongruent, and apolitical headlines.

Results

Truth Sensitivity. A preregistered exploratory 3 (Source Reliability: reliable vs. unreliable vs. no source) $\times$ 3 (Headline Slant: congruent vs. incongruent vs. apolitical) mixed ANOVA with truth sensitivity as the criterion yielded a significant main effect of Source Reliability, $F(2, 873) = 3.89, p = .021, \eta_p^2 = .009, BF_{10} = 0.36$, although Bayes factor analysis suggests little support for the alternative hypothesis. Post-hoc tests with Bonferroni correction indicated that participants in the no-source condition were significantly more truth-sensitive than those in the unreliable-source condition, $p = .032, d = 0.211, BF_{10} = 7.84$ (see Table 4). Truth sensitivity did not significantly differ between the no-source and reliable-source conditions, $p = 1.00, d = 0.025, BF_{10} = 0.06$. The difference between the reliable-source and unreliable-source conditions did not reach statistical significance, $p = .074, d = 0.186, BF_{10} = 2.70$. The main effect of Headline Slant was statistically significant, $F(2, 1746) = 106.00, p < .001, \eta_p^2 = .108, BF_{10} = 1.13 \times 10^{14}$, with participants being significantly better at discerning true from false information when the information was apolitical than when it was ideology-congruent, $p < .001, dz = 0.449, BF_{10} = 5.30 \times 10^{40}$, and ideology-incongruent, $p < .001, dz = 0.416, BF_{10} = 3.12 \times 10^{35}$ (see Table 4). Truth sensitivity did not significantly differ for ideology-congruent and ideology-incongruent headlines, $p = 1.00, dz = 0.029, BF_{10} = 0.06$. The interaction between Headline Slant and Source Reliability was not statistically significant, $F(4, 1746) = .46, p = .762, \eta_p^2 = .001, BF_{10} = 0.00$.

Acceptance Threshold. The preregistered confirmatory 3 (Source Reliability: reliable vs. unreliable vs. no source) $\times$ 3 (Headline Slant: congruent vs. incongruent vs. apolitical) mixed ANOVA with acceptance threshold as the criterion (see Figure 5) yielded no significant main effect of Source Reliability, $F(2, 873) = 2.54, p = .080, \eta_p^2 = .006, BF_{10} = 0.14$. Post-hoc tests

with Bonferroni correction revealed a difference between the reliable-source and unreliable-source conditions in the expected direction, but this difference was not statistically significant, $p = .097$, $d = 0.178$, $BF_{10} = 2.80$. The difference between the reliable-source and no-source condition, $p = .287$, $d = 0.138$, $BF_{10} = 0.67$, and the unreliable-source and no-source condition, $p = 1.000$, $d = 0.040$, $BF_{10} = 0.07$, were not significant.

The main effect of Headline Slant was statistically significant, $F(1.73, 1507.95) = 429.18$, $p < .001$, $\eta_p^2 = .330$, $BF_{10} > 10^{100}$, with Greenhouse-Geisser correction applied to address non-sphericity. Acceptance thresholds for ideology-congruent headlines were significantly lower than for ideology-incongruent headlines, $p < .001$, $d = 1.415$, $BF_{10} = 2.98 \times 10^{95}$, and apolitical headlines, $p < .001$, $d = 0.486$, $BF_{10} = 8.71 \times 10^{12}$, while acceptance thresholds for ideology-incongruent headlines were significantly higher than for apolitical headlines, $p < .001$, $d = 0.822$, $BF_{10} = 1.80 \times 10^{96}$. The interaction between Headline Slant and Source Reliability was not statistically significant, $F(3.46, 1507.95) = 0.28$, $p = .867$, $\eta_p^2 = .001$, $BF_{10} = 1.00$.

Discussion

The results of Experiment 3 support Hypothesis 1, in that participants showed significantly lower acceptance thresholds for ideology-congruent than ideology-incongruent headlines (i.e., partisan bias). However, counter to Hypothesis 2, the effect of source reliability on acceptance thresholds was not significantly moderated by headline type. The latter finding conflicts with the idea that partisan bias nullifies the predicted effect of source reliability on acceptance thresholds for politically charged information. If that were the case, source reliability should produce the expected effects on acceptance thresholds for apolitical information, but not for politically charged information. The results of Experiment 3 do not support this hypothesis.

In addition to these findings, Experiment 3 obtained two unanticipated effects on truth discernment. First, participants showed greater accuracy in discerning true from false information when the information was presented with no source as compared to when the information was presented with an unreliable source. Because the effect did not emerge in Experiments 2a and 2b, we refrain from interpreting this inconsistent finding. Second, participants showed greater accuracy in discerning true from false information when the information was apolitical than when the information was politically charged. Because we did not match the political and apolitical headlines in terms of task-relevant properties, differences in participants' truth discernment for political versus apolitical headlines could potentially stem from stimulus effects, such that greater accuracy in discerning the truth of apolitical information than political information may be explained by the former set of headlines being relatively easier to judge than the latter. It would thus be premature to interpret this finding without further experiments that control for the baseline discriminability of the political and apolitical information.

Experiment 4

The findings presented thus far fail to paint a coherent picture about effects of source reliability. Although source reliability influenced truth judgments in the predicted manner in Experiment 1, source reliability showed inconsistent effects in the following experiments, and none of these effects support our hypothesis that acceptance thresholds would be lower for reliable compared to unreliable sources. In trying to reconcile the conflicting findings, we identified one important difference between the reported studies. While participants in Experiment 1 judged headlines from two different sources (i.e., within-subjects manipulation of source reliability), participants in the subsequent experiments judged headlines from only a

single source (i.e., between-subjects manipulation of source reliability). Whereas the former scenario may foster comparisons between sources, comparative thoughts about source characteristics may be less likely in the latter scenario. Experiment 4 tested if the expected effects of source reliability on acceptance thresholds emerge when participants judge headlines from two different sources, as in Experiment 1. Toward this end, all participants in Experiment 4 judged both headlines paired with a reliable source and headlines paired with an unreliable source (instead of judging headlines paired with only a single source). Critically, the headlines were not paired with the sources in a way that would re-introduce the confound between source reliability and information veracity. Instead, orthogonal to the manipulation of information veracity and headline slant, half of the headlines were presented with a reliable source and half were presented with an unreliable source.

Consistent with the partisan bias found in our earlier experiments, we expected that acceptance thresholds will be significantly lower for ideology-congruent than ideology-incongruent information (Hypothesis 1). Moreover, consistent with the predictions laid out in the lower section of Table 1, we hypothesized that acceptance thresholds will significantly differ across source conditions, such that acceptance thresholds will be significantly lower for information paired with the reliable source than information paired with the unreliable source (Hypothesis 2). Like Experiment 3, analyses of truth sensitivity were conducted exploratorily in Experiment 4 to assess whether source information would have an impact on truth sensitivity in a fully crossed 2 (Accuracy: true vs. false) $\times$ 2 (Source: reliable vs. unreliable) design. The design, hypotheses, and analysis plan of Experiment 4 were preregistered prior to data collection at <https://osf.io/nhtd8/>. Experiment 4's data were collected in December 2023.

Method

Design. Experiment 4 employed a 2 (Headline Accuracy: true vs. false) $\times$ 2 (Headline Slant: pro-Democrat vs. pro-Republican) $\times$ 2 (Source Reliability: reliable vs. unreliable) $\times$ 2 (Political Affiliation: Democrat vs. Republican) mixed design, with the first three factors varying within-subjects and the last factor varying between-subjects. Unlike the earlier experiments in which sources varied between-subjects, all participants judged 40 headlines paired with a reliable source (i.e., Forbes) and 40 headlines paired with an unreliable source (i.e., National Enquirer). Among the 40 headlines paired with each source, the headlines varied in terms of their slant and veracity (i.e., 10 true pro-Democrat, 10 false pro-Democrat, 10 true pro-Republican, 10 false pro-Republican). To avoid item effects, we created two stimulus sets, such that the 40 headlines paired with a reliable source in the first stimulus set would be paired with an unreliable source in the second stimulus set, and vice versa. In other words, all participants judged the same 80 headlines. The order of the 80 headlines was randomized individually for each participant.

Participants. Participants were recruited using the same procedures as the earlier experiments. We aimed for a sample of 400 participants (200 per political affiliation) prior to exclusions. Anticipating that approximately 15% of the participants would be excluded based on preregistered exclusion criteria, a sample of 340 after exclusions provides 80% power for the detection of a small within-subjects effect of $f = 0.09$ in a 2 (Source Reliability: reliable vs. unreliable) $\times$ 2 (Headline Slant: ideology-congruent vs. ideology-incongruent) ANOVA for repeated measures on both factors, assuming a correlation between measures of $r = .30$ and using a nonsphericity correction of $\epsilon = 1$. Participants who completed the study were compensated US-\$3.00.

Four-hundred-and-four participants started the study. All passed a bot-check but four did not complete the study, leaving us with 400 participants who were compensated (200 per political affiliation). Eighteen failed the attention check and another 10 provided inconsistent information on their political affiliation. The final sample was thus $N = 372$. Table 2 provides information on participant demographics.

Measures. Participants judged 80 headlines that differed in terms of their veracity and political slant. The headlines comprised those tested for Experiment 3 and those pre-tested by Gawronski et al. (2023). We created two stimulus sets of headlines. In the first stimulus set, 40 headlines (i.e., 10 true pro-Democrat, 10 false pro-Democrat, 10 true pro-Republican, 10 false pro-Republican) were paired with a reliable source and the other 40 headlines (i.e., 10 true pro-Democrat, 10 false pro-Democrat, 10 true pro-Republican, 10 false pro-Republican) were paired with an unreliable source. In the second stimulus set, the source-headline pairings from the first set of headlines were flipped.

Procedure. The procedures of Experiment 4 largely followed the earlier studies, with the exception that participants were not randomly assigned to see different source information. Instead, participants were randomly assigned to one of the two stimulus set conditions. The order of the 80 headlines was randomized individually for each participant. The study took approximately 15 minutes to complete.

Data Analysis. Data exclusions were identical to the earlier experiments. Like in the first three experiments, SDT indices were calculated separately for ideology-congruent and ideology-incongruent headlines.

Results

Truth Sensitivity. A preregistered exploratory 2 (Source Reliability: reliable vs. unreliable) $\times$ 2 (Headline Slant: ideology-congruent vs. ideology-incongruent) repeated-measures ANOVA with truth sensitivity as the criterion yielded no significant main effect of Headline Slant, $F(1, 371) = 3.41, p = .066, \eta_p^2 = .009, BF_{10} = 0.27$. The main effect of Source Reliability and its interaction with Headline Slant were also not statistically significant, $F_s < 1.33, p_s > .250, \eta_p^2_s < .004, BF_{10s} < 0.08$ (see Table 4).

Acceptance Threshold. The preregistered confirmatory 2 (Source Reliability: reliable vs. unreliable) $\times$ 2 (Headline Slant: ideology-congruent vs. ideology-incongruent) repeated-measures ANOVA with acceptance thresholds as the criterion (see Figure 6) yielded a significant main effect of Headline Slant, $F(1, 371) = 311.10, p < .001, \eta_p^2 = .456, BF_{10} > 10^{100}$, with participants showing significantly lower acceptance thresholds for ideology-congruent information than ideology-incongruent information. The main effect of Source Reliability also reached statistical significance, $F(1, 371) = 12.66, p < .001, \eta_p^2 = .033, BF_{10} = 25.07$, with participants having lower acceptance thresholds for information paired with a reliable source than information paired with an unreliable source. The interaction between Source Reliability and Headline Slant was not statistically significant, $F(1, 371) = 0.09, p = .765, \eta_p^2 < .001, BF_{10} = 0.33$.

Discussion

Supporting Hypothesis 1, acceptance thresholds were again lower for ideology-congruent information compared to ideology-incongruent information. More importantly, varying source reliability within-participants, acceptance thresholds were lower for information presented with a reliable source as compared to information presented with an unreliable source, supporting

Hypothesis 2. These results are consistent with the idea that the lack of source-reliability effects in Experiments 2 and 3 is due to the lack of comparative thinking when source reliability is manipulated between-subjects. Moreover, in a fully crossed 2 (Accuracy: true vs. false) $\times$ 2 (Source: reliable vs. unreliable) within-subjects design that deconfounds effects of source reliability from effects of information veracity, source reliability affects truth judgments via acceptance thresholds rather than truth sensitivity. Thus, the identified effect of source information on truth sensitivity in Experiment 1 can be described as a spurious product of the incomplete design whose true nature is revealed when all cases are considered (see Table 1).

Experiment 5

Experiment 5 investigated effects of a second source characteristic: perceived partisanship. A comparison of different source characteristics is important because perceived reliability and perceived partisanship may show distinct effects. Whereas perceived reliability may influence acceptance thresholds irrespective of the information content, effects of perceived partisanship may depend on whether the political slant of the information aligns with the recipient's political views. In line with the principles of augmentation and discounting in classic attribution theory (Kelley & Michela, 1980), acceptance thresholds for ideology-incongruent information may be lower when the information comes from an ideology-congruent source (e.g., a Democrat seeing pro-Republican information on CNN; a Republican seeing pro-Democrat information on Fox News) than when it comes from an ideology-incongruent source (e.g., a Democrat seeing pro-Republican information on Fox News; a Republican seeing pro-Democrat information on CNN). Conversely, for ideology-congruent information, acceptance thresholds may be lower when the information comes from an ideology-incongruent source (e.g., a Democrat seeing pro-Democrat information reported on Fox News; a Republican seeing pro-

Republican information reported on CNN) than when it comes from an ideology-congruent source (e.g., a Democrat seeing pro-Democrat information reported on CNN; a Republican seeing pro-Republican reported on Fox News). Moreover, because perceived partisanship involves expectancies that are specific to political information, perceived partisanship may have no effect on truth judgments of apolitical information. The main goal of Experiment 5 was to test these predictions.

To this end, Experiment 5 used the design of Experiment 3, the only difference being that we manipulated perceived source partisanship rather than perceived source reliability. Specifically, instead of presenting all headlines with either (1) a reliable non-partisan source, (2) an unreliable non-partisan source, or (3) no source, we presented all headlines with either (1) a pro-Republican source (i.e., Fox News), (2) a pro-Democrat source (i.e., CNN), or (3) no source. As in Experiment 3, the stimuli included headlines with a pro-Republican slant, headlines with a pro-Democrat slant, and apolitical headlines. Headlines and sources were both coded in terms of whether they are congruent or incongruent with participants' political leaning (or apolitical).

Consistent with the results of our earlier studies, we hypothesized that acceptance thresholds will differ across headline types (Hypothesis 1a), such that acceptance thresholds would be lower for ideology-congruent than ideology-incongruent headlines (Hypothesis 1b). More important for the current question, we hypothesized that the effect of source partisanship would depend on headline type (Hypothesis 2a). For ideology-congruent headlines, we tested the hypotheses that acceptance thresholds would be higher in the ideology-congruent source condition than the no-source condition (Hypothesis 2b), higher in the no-source condition than the ideology-incongruent source condition (Hypothesis 2c), and higher in the ideology-congruent source than the ideology-incongruent condition (Hypothesis 2d). For ideology-incongruent

headlines, we tested the hypotheses that acceptance thresholds would be lower in the ideology-congruent source condition than the no-source condition (Hypothesis 2e), lower in the no-source condition than the ideology-incongruent source condition (Hypothesis 2f), and lower in the ideology-congruent source condition than the ideology-incongruent source condition (Hypothesis 2g). For apolitical headlines, we hypothesized that acceptance thresholds would not differ between the ideology-congruent source condition and the no-source condition (Hypothesis 2h), between the no-source condition and the ideology-incongruent source condition (Hypothesis 2i), and between the ideology-congruent source condition and the ideology-incongruent source condition (Hypothesis 2j). The design, hypotheses, and analysis plan of Experiment 5 were preregistered prior to data collection at <https://osf.io/2dk4v/>. Experiment 5's data were collected in May 2024.

Method

Design. Experiment 5 employed a 2 (Headline Accuracy: true vs. false) $\times$ 3 (Headline Slant: pro-Democrat vs. pro-Republican vs. apolitical) $\times$ 2 (Political Affiliation: Democrat vs. Republican) $\times$ 3 (Source Partisanship: pro-Democrat vs. pro-Republican vs. no source) mixed design, with the first two factors varying within-subjects and the latter two factors varying between-subjects. Participants were randomly assigned to judge all headlines presented with either a pro-Republican source (Fox News), a pro-Democrat source (CNN), or no source.

Participants. We aimed for a sample of 900 participants prior to exclusions (450 per political affiliation). Anticipating that approximately 15% of the participants would be excluded based on preregistered exclusion criteria, a sample of 765 after exclusions provides 80% power for the detection of a small interaction effect of $f = 0.060$ in a 3 (Source Partisanship: congruent-source vs. incongruent-source vs. no source) $\times$ 3 (Headline Slant: ideology-congruent vs.

ideology-incongruent vs. apolitical) mixed ANOVA with the first variable as a between-subjects factor and the second as a within-subjects factor, assuming a correlation between measures of $r = .30$ and using a nonsphericity correction of $\epsilon = 1$. Participants who completed the study were compensated US-\$3.00.

Nine-hundred-and-sixteen participants started the study. One failed to pass a bot-check and another 14 did not complete the entire study. The final sample included 901 participants with valid responses. We had stopped data collection once we compensated 900 participants. One additional participant completed all items, did not click the final submit button, and did not request compensation. Of the 901 participants, 30 failed the attention check and another nine provided inconsistent information on their political affiliation. The final sample was thus $N = 862$. Table 2 provides information on participant demographics.

Measures. The 60 headlines included in the study varied in terms of whether they are (1) true or false and (2) pro-Democrat, pro-Republican, or apolitical, with each of the six categories having 10 headlines. The headlines used in Experiment 5 comprised a mix of headlines tested for the earlier experiments reported here and those pre-tested by Gawronski et al. (2023).

Procedure. The procedures of Experiment 5 followed that of Experiment 3, the only difference being that participants were randomly assigned to judge all 60 headlines paired with either (1) a pro-Democrat source (CNN), (2) a pro-Republican source (Fox News), or (3) no source, instead of reliable and unreliable non-partisan sources. The order of the 60 headlines was randomized individually for each participant. The study took approximately 15 minutes to complete.

Data Analysis. Data exclusions and analytic approaches followed that of Experiment 3, with d' and c scores calculated separately for ideology-congruent, ideology-incongruent, and apolitical headlines.

Results

Truth Sensitivity. A preregistered exploratory 3 (Source Partisanship: congruent-source vs. incongruent-source vs. no source) $\times$ 3 (Headline Slant: ideology-congruent vs. ideology-incongruent vs. apolitical) mixed ANOVA with truth sensitivity as the criterion yielded no significant main effect of Source Partisanship, $F(2, 859) = 0.85, p = .427, \eta_p^2 = .002, BF_{10} = 0.02$. The main effect of Headline Slant was statistically significant, $F(1.98, 1699.24) = 29.90, p < .001, \eta_p^2 = .034, BF_{10} = 1.32 \times 10^{10}$, with Greenhouse-Geisser correction applied to address non-sphericity (see Table 4). Participants were significantly better at discerning true from false information when the information was apolitical than when it was ideology-congruent, $p < .001, d = 0.403, BF_{10} = 3.98 \times 10^8$, and ideology-incongruent, $p < .001, d = 0.401, BF_{10} = 2.99 \times 10^8$ (see Table 4). Truth sensitivity did not significantly differ for ideology-congruent and ideology-incongruent headlines, $p = 1.00, d = 0.008, BF_{10} = 0.04$. The interaction between Headline Slant and Source Partisanship was not statistically significant, $F(3.96, 1699.24) = 0.99, p = .410, \eta_p^2 = .002, BF_{10} = 0.00$.

Acceptance Threshold. The preregistered confirmatory 3 (Source Partisanship: congruent-source vs. incongruent-source vs. no-source) $\times$ 3 (Headline Slant: ideology-congruent vs. ideology-incongruent vs. apolitical) mixed ANOVA with acceptance threshold as the criterion (see Figure 7) yielded no significant main effect of Source Partisanship, $F(2, 859) = 0.23, p = .797, \eta_p^2 = .001, BF_{10} = 0.01$. The main effect of Headline Slant was statistically significant, $F(1.75, 1505.64) = 426.20, p < .001, \eta_p^2 = .332, BF_{10} > 10^{100}$, with Greenhouse-Geisser

correction applied to address non-sphericity. Acceptance thresholds for ideology-congruent headlines were significantly lower than for ideology-incongruent headlines, $p < .001$, $d = 1.458$, $BF_{10} = 5.86 \times 10^{99}$, and apolitical headlines, $p < .001$, $d = 0.810$, $BF_{10} = 1.02 \times 10^{36}$, while acceptance thresholds for ideology-incongruent headlines were significantly higher than for apolitical headlines, $p < .001$, $d = 1.090$, $BF_{10} = 7.22 \times 10^{61}$. The interaction between Headline Slant and Source Partisanship was not statistically significant, $F(3.46, 1507.95) = 0.28$, $p = .867$, $\eta_p^2 = .001$, $BF_{10} = 0.00$.

Discussion

The results of Experiment 5 support Hypotheses 1a and 1b, in that acceptance thresholds differed across headline types, with acceptance thresholds being lower for ideology-congruent than ideology-incongruent headlines (i.e., partisan bias; see Batailler et al., 2022, Gawronski et al., 2023). Although not the focus of the current work, truth sensitivity was greatest when the information was politically neutral than when the information was politically charged, replicating an unexpected finding of Experiment 3. As aforementioned in the discussion of Experiment 3's findings, the political and apolitical headlines were not matched in terms of task-relevant properties. Thus, the finding that participants showed better truth discernment for apolitical information than for political information could potentially stem from stimulus effects, and we therefore refrain from interpreting this effect. More importantly, source partisanship did not affect acceptance thresholds. Neither the main effect of source partisanship nor its interaction with headline type reached statistical significance. Thus, our hypothesis that effects of source partisanship depend on information content was not supported.

Experiment 6

Because the predicted effects of source reliability in our earlier experiments were limited to conditions when participants judged information from more than one source (presumably promoting comparisons between different sources), we proceeded to test effects of source partisanship using the within-subjects design of Experiment 4. Based on the strong partisan bias found in our earlier experiments, we hypothesized that acceptance thresholds will be lower for ideology-congruent information than ideology-incongruent information (Hypothesis 1). We also tested the preregistered hypotheses that the effect of source partisanship will depend on headline type (Hypothesis 2a), such that acceptance thresholds for ideology-congruent headlines will be lower when the source is ideologically incongruent than when the source is ideologically congruent (Hypothesis 2b), and that acceptance thresholds for ideology-incongruent headlines will be higher when the source is ideologically incongruent than when the source is ideologically congruent (Hypothesis 2c). The design, hypotheses, and analysis plan of Experiment 6 were preregistered prior to data collection at <https://osf.io/b6uj2/>. Experiment 6's data were collected in May 2024.

Method

Design. Experiment 6 employed a 2 (Headline Accuracy: true vs. false) $\times$ 2 (Headline Slant: pro-Democrat vs. pro-Republican) $\times$ 2 (Source Partisanship: pro-Republican vs. pro-Democrat) $\times$ 2 (Political Affiliation: Democrat vs. Republican) mixed design, with the first three factors varying within-subjects and the last factor varying between-subjects. All participants judged 80 headlines, 40 of which were paired with a pro-Republican source (i.e., Fox News) and 40 of which were paired with a pro-Democrat source (i.e., CNN). Among the 40 headlines paired with each source, the headlines varied in terms of their slant and veracity (i.e., 10 true pro-

Democrat, 10 false pro-Democrat, 10 true pro-Republican, 10 false pro-Republican). To avoid item effects, we created two stimulus sets, such that the 40 headlines paired with a pro-Republican source in the first stimulus set would be paired with a pro-Democrat source in the second stimulus set, or vice versa. The order of the 80 headlines was randomized individually for each participant.

Participants. Participants were recruited using the same procedures as in the earlier experiments. We aimed for a sample of 400 participants (200 per political affiliation) prior to exclusions. Anticipating that approximately 15% of the participants would be excluded based on preregistered exclusion criteria, a sample of 340 after exclusions provides 80% power for the detection of a small within-subjects effect of $f = 0.09$ in a 2 (Source Partisanship: congruent-source vs. incongruent-source) $\times$ 2 (Headline Slant: ideology-congruent vs. ideology-incongruent) ANOVA for repeated measures on both factors, assuming a correlation between measures of $r = .30$ and using a nonsphericity correction of $\epsilon = 1$. Participants who completed the study were compensated US-\$3.00.

Four-hundred-and-seven participants started the study. All passed a bot-check but seven did not complete the study, leaving us with 400 participants who were compensated (200 per political affiliation). Sixteen failed the attention check and another six provided inconsistent information on their political affiliation. The final sample was thus $N = 378$. Table 2 provides information on participant demographics.

Measures. Participants judged 80 headlines that differed in terms of their veracity and political slant. The headlines used in Experiment 6 comprised a mix of headlines tested for the earlier experiments reported here and those pre-tested by Gawronski et al. (2023). We created two stimulus sets of headlines. In the first stimulus set, 40 headlines (i.e., 10 true pro-Democrat,

10 false pro-Democrat, 10 true pro-Republican, 10 false pro-Republican) were paired with a pro-Republican source and the other 40 headlines (i.e., 10 true pro-Democrat, 10 false pro-Democrat, 10 true pro-Republican, 10 false pro-Republican) were paired with a pro-Democrat source. In the second stimulus set, the source-and-headline-slant pairings from the first set of headlines were flipped.

Procedure. The procedures of Experiment 6 followed that of Experiment 4, with participants randomly assigned to one of the two stimulus set conditions. The order of the 80 headlines was randomized individually for each participant. The study took approximately 15 minutes to complete.

Data Analysis. Data exclusions and calculation of SDT indices proceeded in accordance with the earlier experiments.

Results

Truth Sensitivity. A preregistered exploratory 2 (Source Partisanship: congruent-source vs. incongruent-source) $\times$ 2 (Headline Slant: ideology-congruent vs. ideology-incongruent) repeated-measures ANOVA with truth sensitivity as the criterion yielded no significant effect of Headline Slant, $F(1, 377) = 1.20, p = .273, \eta_p^2 = .003, BF_{10} = 0.09$. The main effect of Source Partisanship, $F(1, 377) = 0.19, p = .665, \eta_p^2 = .000, BF_{10} = 0.04$, and the interaction between Headline Slant and Source Partisanship were also not statistically significant, $F(1, 377) = 0.30, p = .584, \eta_p^2 = .001, BF_{10} = 0.00$ (see Table 4).

Acceptance Threshold. The preregistered confirmatory 2 (Source Partisanship: congruent-source vs. incongruent-source) $\times$ 2 (Headline Slant: ideology-congruent vs. ideology-incongruent) repeated-measures ANOVA with acceptance thresholds as the criterion (see Figure 8) yielded a significant main effect of Headline Slant, $F(1, 377) = 298.92, p < .001, \eta_p^2 = .442$,

$BF_{10} > 10^{100}$, with participants showing significantly lower acceptance thresholds for ideology-congruent information than ideology-incongruent information. The main effect of Source Partisanship also reached statistical significance, $F(1, 377) = 12.60$ $p < .001$, $\eta_p^2 = .032$, $BF_{10} = 10.89$, with participants having lower acceptance thresholds for information paired with an ideology-congruent source than information paired with an ideology-incongruent source. The interaction term was not statistically significant, $F(1, 377) = 0.01$, $p = .916$, $\eta_p^2 < .001$, $BF_{10} = 0.29$.

Discussion

Consistent with the partisan bias found in our earlier experiments, acceptance thresholds were significantly lower for ideology-congruent information compared to ideology-incongruent information. In addition, consistent with the pattern obtained for source reliability in Experiment 4, acceptance thresholds were significantly lower for information presented with an ideology-congruent source than for information presented with an ideology-incongruent source. However, counter to our preregistered hypotheses, the effects of source partisanship did not depend on information content. That is, acceptance thresholds were lower for information presented with an ideology-congruent source than for information presented with an ideology-incongruent source regardless of whether the focal information was ideology-congruent or ideology-incongruent. Thus, across the current series of experiments, source reliability and source partisanship showed the same pattern of results, in that (1) the effects of both source facets generalized across information content (i.e., ideology-congruent vs. ideology-incongruent information), and (2) the effects of both facets were limited to conditions where comparisons of sources are salient (i.e., within-subjects vs. between-subjects manipulation of source information).

General Discussion

Across seven preregistered experiments, the current research adopted a signal-detection approach to investigate: (1) whether source characteristics influence truth judgments of true and false information via truth sensitivity or acceptance thresholds; (2) potential differences between perceived reliability and perceived partisanship as distinct facets of sources; and (3) how the two facets interact with information content in shaping judgments of true and false information. Although several of our preregistered hypotheses were not confirmed, the pattern of results across the seven experiments reveals a highly coherent picture (see Table 5), indicating that (1) source characteristics affected truth judgments of true and false information via acceptance thresholds, not truth sensitivity *per se*; (2) effects of source characteristics were limited to conditions where comparisons of sources are salient; (3) source characteristics had no effect on partisan bias, in that acceptance thresholds were substantially lower for ideology-congruent than ideology-incongruent information regardless of the source; and (4) perceived reliability and perceived partisanship showed corresponding effects on acceptance thresholds, regardless of information content.

Source Effects on Truth Sensitivity

Our findings suggest that the widespread practice of confounding source credibility with information veracity (i.e., by presenting false information with its original dubious source and true information with its original mainstream source; e.g., Pennycook & Rand, 2019) leads to a spurious increase in truth sensitivity. Hence, any identified effect in studies that adopted this approach could be driven by either the perceived credibility of the source or the veracity of the information (or both). When we confounded source reliability and information veracity by presenting true information with a reliable source and false information with an unreliable source

in Experiment 1 (see Table 1, upper section), we found that truth sensitivity was greater when information about the source was present than when it was absent. However, when we deconfounded source reliability from information veracity in a fully crossed design in Experiments 2a–4 (see Table 1, lower section), source reliability did not reliably affect truth sensitivity. Instead, we found that source reliability affected truth judgments via acceptance thresholds, such that participants were more likely to accept information from reliable than unreliable sources irrespective of whether the information was true or false. Although the latter effect depended on the salience of comparisons between sources, these results challenge conclusions from meta-analytic findings that source information increases people’s ability to discern true from false information (see Sultan et al., 2024).

Source Effects on Acceptance Thresholds

The pattern of results across the seven studies suggests that source characteristics influence truth judgments via acceptance thresholds, not truth sensitivity *per se*. That is, when source characteristics and information veracity were manipulated independently in a fully crossed design in Experiments 2a–6, participants were more likely to accept information from reliable and ideology-congruent sources than unreliable and ideology-incongruent sources, irrespective of whether the information was true or false. However, source characteristics influenced acceptance thresholds only when participants judged information from more than one source (i.e., when source characteristics were manipulated within-subjects), but not when participants judged information from a single source (i.e., when source characteristics were manipulated between-subjects). A potential interpretation of this finding is that judging information from more than one source promotes comparative thinking about different sources (i.e., National Enquirer vs. Forbes; CNN vs. Fox News), which may amplify effects of source

characteristics on truth judgments (e.g., salience and contrast effects; Mussweiler, 2003; Taylor & Fiske, 1978; Tversky & Simonson, 1993). Conversely, when judging information from a single source, comparative thinking about source characteristics is likely less pronounced, which may reduce the influence of source characteristics on truth judgments. Yet, even under conditions where source characteristics reliably influenced acceptance thresholds, their impact was relatively small overall, with effect sizes of $\eta_p^2 = .033$ (Experiment 4) and $.032$ (Experiment 6), respectively.

Robustness of Partisan Biases against Source Cues

Across all seven experiments, we found robust evidence for content-driven partisan bias, in that participants showed significantly lower acceptance thresholds for ideology-congruent information than ideology-incongruent information. In terms of current metrics for the interpretation of effect sizes (Cohen, 1988), partisan-bias effects far exceeded conventional benchmarks for a large effect (i.e., $\eta_p^2 = .14$; $d = 0.80$), with effect sizes of $\eta_p^2 \geq .442$ (Experiments 1, 2a, 2b, 4, and 6) and $d \geq 1.458$ (Experiments 3 and 5). Bayes factor analyses also indicated strong support for the alternative hypothesis over the null hypothesis ($BF_{10s} > 4.88 \times 10^{13}$). Importantly, partisan-bias effects were highly robust, in that they remained unaffected by source characteristics across all seven experiments. In other words, participants readily accepted ideology-congruent information even when it came from a source they perceived to be unreliable (i.e., National Enquirer) or a source they perceived to be in opposition to their preferred political party (i.e., Fox News for Democrats; CNN for Republicans). Conversely, participants readily rejected ideology-incongruent information even when it came from a source they perceived to be reliable (i.e., Forbes) or a source they perceived to support their preferred political party (i.e., CNN for Democrats; Fox News for Republicans).

Essentially, the current findings suggest that (1) regardless of whether information veracity was confounded with source reliability or not and (2) regardless of whether participants judged headlines from a single source or multiple sources of varying reliability or partisanship, participants were more likely to accept information that was congruent with their political beliefs as true and to reject information that was incongruent with their political beliefs as false. Importantly, across all seven experiments and conditions, source characteristics did not attenuate a well-documented asymmetry in partisan-bias effects (see Gawronski et al., 2026): while acceptance thresholds for ideology-incongruent information consistently exceeded zero, acceptance thresholds for ideology-congruent information tended to be close to zero. These results suggest that skepticism towards ideology-incongruent information is far more pronounced than gullibility towards ideology-congruent information, given the larger absolute c values of the former compared to the latter across all seven experiments and conditions. This asymmetric partisan-bias effect persisted across all studies regardless of source manipulations.

In addition to providing further evidence for the pervasiveness and robustness of partisan bias in judgments of political (mis)information (Gawronski et al., 2025), the current findings are noteworthy because they suggest that the ideological congruence of the information content is much more influential in shaping truth judgments than characteristics of the source. Specifically, our findings suggest that partisans (at least those in the United States) do not readily accept information from reliable and ideology-congruent sources and that they do not readily reject information from unreliable and ideology-incongruent sources. Instead, they do so in a way that supports and protects their political worldviews (Ditto et al., 2025; Van Bavel et al., 2021, 2024; Van Bavel & Pereira, 2018).

Source Partisanship $\approx$ Source Reliability

Experiments 1 to 4 examined the effects of perceived source reliability of non-partisan sources, whereas Experiments 5 and 6 tested effects of source partisanship. Overall, our findings suggest that the mechanisms through which perceived reliability and perceived partisanship affect truth judgments of political information are functionally similar, in that (1) both source characteristics shape truth judgments via acceptance thresholds rather than truth sensitivity; (2) both source characteristics do not affect partisan bias in judgments of ideology-congruent and ideology-incongruent information; and (3) effects of neither source characteristic depend on information content. These findings point to two possibilities. First, perceived source reliability and perceived source partisanship, while being two distinct source characteristics, may have the same functional effect on truth judgments. Second, the effects of perceived source reliability and perceived source partisanship could be both mediated by overall perceived source credibility. The latter is a possibility in the current research, as the source manipulations in Experiments 5 and 6 did not fully disentangle source reliability from source partisanship, in that the partisan sources also differed in their perceived reliability among the different groups of partisan audiences. In other words, source partisanship may act as a proxy for perceived credibility in the politically-charged media landscape. This conclusion is consistent with findings by Traberg and van der Linden (2022), suggesting that partisans perceive ideology-congruent sources to be less biased and more credible than ideology-incongruent sources.

These findings are noteworthy because effects of perceived source partisanship may be expected to depend on whether the focal information is congruent or incongruent with a person's political leaning. On the one hand, the low subjective likelihood of ideology-congruent information being reported by an ideology-incongruent source may lead people to assume that

the information must be true (e.g., a Democrat seeing pro-Democrat information reported on Fox News; a Republican seeing pro-Republican reported on CNN). On the other hand, the low subjective likelihood of ideology-incongruent information being reported by an ideology-congruent source may similarly lead people to assume that the information must be true (e.g., a Democrat seeing pro-Republican information reported on CNN; a Republican seeing pro-Democrat reported on Fox News). In line with the principles of augmentation and discounting in classic attribution theories (Kelley & Michela, 1980), these considerations would suggest that effects of perceived partisanship may depend on the information content, in that (1) acceptance thresholds for ideology-congruent information should be lower for ideology-incongruent than ideology-congruent sources, and conversely (2) acceptance thresholds for ideology-incongruent information should be lower for ideology-congruent than ideology-incongruent sources. Moreover, because perceived partisanship involves expectancies that are specific to political information, one may expect no effects for apolitical information. Our findings disconfirmed these predictions and instead showed that participants were more likely to accept information from ideology-congruent sources than ideology-incongruent sources, irrespective of whether the focal information was ideology-congruent, ideology-incongruent, or apolitical. This pattern mirrors the one obtained for source reliability, in that participants were more likely to accept information from reliable sources than unreliable sources, irrespective of whether the focal information was ideology-congruent, ideology-incongruent, or apolitical.

That being said, effects of both perceived reliability and perceived partisanship were limited to conditions where comparisons between sources were salient, and the size of source effects under such conditions (η_p^2 s = .033 and .032, respectively) was relatively small in terms of current metrics for the interpretation of effect sizes (Cohen, 1988). The size of these effects also

stands in stark contrast to the extremely large partisan-bias effects of information content (η_p^2 s $\geq .442$; $ds \geq 1.458$), which far exceeded conventional benchmarks for a large effect (Cohen, 1988). Thus, while participants were more likely to trust information from sources they deem credible (vs. non-credible), participants' truth judgments were much more strongly shaped by the political slant of the focal information, in that they accepted ideology-congruent information and rejected ideology-incongruent information regardless of whether they perceived the source to be reliable or unreliable and regardless of whether they perceived the source to be ideology-congruent or ideology-incongruent (and thus credible or non-credible, respectively).

Implications

If people were homo rationales, they would trust reliable, non-partisan sources to convey political information that is factually true, regardless of whether the information challenges or supports their political beliefs. Yet, our findings suggest that, despite both Forbes and National Enquirer being seen as relatively non-partisan, participants rejected political information conveyed by the two sources when this information supported the opposing political group or challenged their own political group. Moreover, our findings suggest that, despite stark differences in perceived reliability between Forbes and National Enquirer (see Table 2), participants were more likely to accept ideology-congruent information from an unreliable, tabloid source like National Enquirer than to accept ideology-incongruent information from a reliable, mainstream source like Forbes. While prior research (Traberg & van der Linden, 2022) and Experiments 5 and 6 of the current research suggest that people tend to fall for misinformation presented by an ideology-congruent source, we demonstrate that even reliable non-partisan sources like Forbes fail to mitigate the large and persistent partisan-bias effects in judgments of political information. Essentially, if the relative influence of the ideological

alignment of the information content outweighs the relative influence of the perceived credibility of the information source (e.g., as signaled by perceived source reliability and source partisanship), interventions that focus on promoting reliance on reliable, politically neutral sources (e.g., NewsGuard web extension; see Acerbi et al., 2022) may fail to mitigate partisan biases in truth judgments of political information.

Open Questions and Limitations

While the current research tested effects of perceived reliability and perceived partisanship among sources that formally operate as newsrooms, we did not examine effects of fake-news sources. Even though the National Enquirer is a tabloid source that is widely perceived as unreliable, it is distinct from fake-news sources, which primarily and deliberately publish hoaxes and disinformation (Allcott & Gentzkow, 2017; Lazer et al., 2018). Fake-news sources vary from using source names that are obviously dubious (e.g., *100PercentFedUp.com* and *Endingthefed.com*) to names mimicking mainstream sources (e.g., *Now8News.com* instead of the legitimate *8NewsNow.com*). It currently remains unclear whether people's biased tendency to accept false information that is ideology-congruent would generalize to fake-news sources, particularly when the sources are obviously dubious. On the one hand, given the findings of past research on source effects, it is possible that people's acceptance thresholds for judging information as true would be particularly high when the sources are obviously lacking credibility. On the other hand, given the large partisan-bias effects detected in the current research (even for an unreliable tabloid news source like the National Enquirer), it is possible that people display the same systematic biases in judgments of information from fake-news sources, albeit potentially to a lesser degree than when the sources are not as obviously questionable. Future research may help to address these questions.

Moreover, while the current research has solely focused on investigating the effects of the creator of the content (i.e., the news sources), future work should also examine effects of the credibility of the source medium (i.e., channel through which the content is conveyed; e.g., social media platforms), the original information source (i.e., the individual or entity providing the information conveyed in the news content; e.g., the politicians quoted in a news story), and the messenger (i.e., the individual or entity who shares the content; e.g., politician vs. friends). Besides focusing solely on effects of mainstream news sources, the current research examined judgments of unambiguously true versus false information; it did not examine judgments of factually accurate but misleading information—another type of content that some researchers classify as misinformation (e.g., van der Linden et al., 2026). Mainstream news sources such as Fox News and CNN seem more likely to publish accurate news content that is presented in a misleading way (e.g., headlines lacking important contextual information) than to publish blatant falsehoods. The higher prevalence of misleading information in the information ecosystem renders it potentially more influential in shaping behavior than plain falsehoods (Allen et al., 2024). Given that people are more likely to encounter misleading information than plain falsehoods from mainstream news sources, examining source effects on judgments of such information is a pertinent task for future research.

It is also worth noting that, while much research on source effects on misinformation susceptibility has focused on a one-directional relationship wherein cues about a source's characteristics are presumed to impact judgments of the information's veracity, extant research has yet to test for the opposite relationship. Specifically, the congruence between the information conveyed and one's political beliefs could potentially impact perceptions of the source itself. Such a reverse effect may emerge when news sources, either real or fake, are unfamiliar.

Considering the vast number of news sources in the United States alone (Institute for Quantitative Social Science, 2021), it seems possible that some news sources provide little to no signal about whether the information one receives is trustworthy or not. The current research design, which intentionally dissociated source characteristics from information veracity, is not designed to test learning effects whereby participants may learn or update their beliefs about a given source's reliability or partisanship across trials. Investigating such learning effects, whereby people's perceptions of a source may change depending on the information they receive, represents an interesting question for future research.

One additional point of consideration is that we only tested the effects of news headlines, but not news content (i.e., an entire article), and source information was only provided in the form of the source's logo in the current studies. Saliency and potency of source effects could vary depending on the way the stimuli were presented (Aslett et al., 2024). For example, when people are presented with full news articles, source characteristics may exert greater influence when other cues signaling credibility (e.g., professionally designed website vs. websites with click-bait advertisements) are present as compared to when the headlines are only presented with the source logo (as in the current studies).

We also acknowledge some limitations that pertain to the context and source manipulations used in the current studies. First, we focused specifically on testing the effects of source characteristics on truth judgments of political information in the United States, which is experiencing high levels of political polarization (Finkel et al., 2020). Whether perceived reliability and perceived partisanship would have a stronger impact on judgments of political information in less polarized contexts remains to be examined.

Second, the current experiments used a single source that met the required criterion. That is, in Experiments 1 to 4, the only reliable non-partisan source that was tagged to the headlines was Forbes, whereas the only unreliable non-partisan source participants saw was National Enquirer. Likewise, in Experiments 5 and 6, we only used CNN as the pro-Democrat source and Fox News as the pro-Republican source. We made this decision based on practical constraints: while there were several sources that met our criterion for inclusion as reliable, non-partisan sources (e.g., AP News), the only source that met our criterion for inclusion as an unreliable, non-partisan source was National Enquirer. Using a single source per source category is limiting, because participants' responses could have been influenced not only by the source's perceived reliability or perceived partisanship, but also by idiosyncratic features of the specific sources. That being said, the fact that the two pairs of sources (i.e., Forbes vs. National Enquirer; CNN vs. Fox News) produced the same pattern of results suggests that our findings reflect general effects that are independent of idiosyncratic features of the chosen sources. Future research could include multiple sources per source category to ensure that the observed effects are wholly attributable to the investigated source characteristics, rather than other source-specific characteristics.

Third, we acknowledge that an empirical test of the effects of source partisanship ideally requires partisan sources that are (1) similarly familiar and (2) similarly reliable to both Democrats and Republicans. In other words, to fully isolate the effects of source partisanship (e.g., Wallace et al., 2020), one would need to identify (1) a source that both Democrats and Republicans perceive to be pro-Democrat, similarly familiar, and similarly reliable, and (2) a source that both Democrats and Republicans perceive to be pro-Republican, similarly familiar, and similarly reliable. However, considering the results of our pilot study (see Supplementary

Materials), this challenge seems insurmountable, given that partisans appear to use source congruence as a cue for source reliability (consistent with the findings of the main experiments). Essentially, pro-Democrat sources were by and large perceived to be unreliable by Republicans whereas pro-Republican sources were by and large perceived to be unreliable by Democrats. We were unable to find a pair of pro-Democrat and pro-Republican sources that are perceived to be similarly reliable to both Democrats and Republicans. Likewise, we were unable to find a pair of pro-Democrat and pro-Republican sources that are similarly unreliable to both Democrats and Republicans. Nevertheless, using CNN and Fox News allowed us to identify effects of partisanship above and beyond the effects of source reliability, and our findings suggest that the perceived partisanship of sources is used as a proxy for source reliability.

Fourth, we acknowledge that pairing true headlines with unreliable sources and false headlines with reliable sources departs from the base-rates in actual information environments. In the real world, reliable news sources are more likely to report true information than unreliable sources, and unreliable news sources are more likely to report false information than reliable sources. The same critique applies to our pairing of pro-Democrat headlines with Fox News and pro-Republican headlines with CNN, in that Fox News is more likely to report pro-Republican news than CNN, whereas CNN is more likely to report pro-Democrat news than Fox News. One might therefore question the generalizability of the present findings to information environments wherein source-information pairings are correlated rather than orthogonal. The current research's pairing of (a) true information with unreliable sources and false information with reliable sources and (b) pro-Democrat information with pro-Republican sources and pro-Republican information with pro-Democrat was necessary to address our theoretical question. Without a fully orthogonal manipulation of source characteristics and information veracity, any observed effects of source

characteristics on truth judgment would be inherently confounded with differences in information veracity. Accordingly, the departures from real-world base rates reflect a necessary methodological choice to isolate the effects of source characteristics (i.e., our focal theoretical question) and thus achieve internal validity, rather than a limitation of the research design.

Conclusion

In sum, the current research aimed to address three questions: (1) Do source characteristics influence truth judgments of true and false information via truth sensitivity or acceptance thresholds? (2) Are there differences in the effects of perceived reliability and perceived partisanship on truth judgments of true and false information? (3) How do the two facets of sources interact with information content in shaping truth judgments of true and false information?

First, we established that source characteristics—operationalized as perceived reliability and perceived partisanship—affects truth judgments through acceptance thresholds, and not truth sensitivity *per se*. Second, the effects of source characteristics are notably more potent when comparisons between different sources were salient. Third, the ideological congruence of the information weighs much more heavily in truth judgments of political information than source characteristics. Fourth, while people are more willing to accept information as true if it is conveyed by a reliable (vs. unreliable) or ideology-congruent (vs. ideology-incongruent) source, they still exhibit strong partisan bias in their judgments by being more accepting of information that is congruent with their political views and being more skeptical of information that is incongruent with their political views. Fifth, source characteristics do not seem to mitigate partisan bias. Sixth, source partisanship does not seem to have any effects on truth judgments that is distinct from source reliability.

Importantly, while biases in media reporting have received much attention in recent years, the current research suggests that people do not readily accept all information from ideology-congruent sources, and they do not readily reject all information from ideology-incongruent sources. Instead, they selectively accept information that supports their political views, and they selectively reject information that conflicts with their political views. The current research also highlights a troubling problem: even when news sources are not biased, people's judgments of information remain so—to the point where unreliable sources are trusted if they affirm one's beliefs and reliable sources are doubted if they challenge one's beliefs.

References

- Acerbi, A., Altay, S., & Mercier, H. (2022). Fighting misinformation or fighting for information? *Harvard Kennedy School Misinformation Review*. <https://doi.org/10.37016/mr-2020-87>
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211–236. <https://doi.org/10.1257/jep.31.2.211>
- Allen, J., Watts, D. J., & Rand, D. G. (2024). Quantifying the impact of misinformation and vaccine-skeptical content on Facebook. *Science*, 384(6699), eadk3451. <https://doi.org/10.1126/science.adk3451>
- Aslett, K., Sanderson, Z., Godel, W., Persily, N., Nagler, J., Bonneau, R., & Tucker, J. A. (2024). Testing the effect of information on discerning the veracity of news in real time. *Journal of Experimental Political Science*, 11(3), 262–276. <https://doi.org/10.1017/XPS.2023.20>
- Batailler, C., Brannon, S. M., Teas, P. E., & Gawronski, B. (2022). A signal detection approach to understanding the identification of fake news. *Perspectives on Psychological Science*, 17(1), 78–98. <https://doi.org/10.1177/1745691620986135>
- Bauer, P. C., & Clemm Von Hohenberg, B. (2021). Believing and sharing information by fake sources: An experiment. *Political Communication*, 38(6), 647–671. <https://doi.org/10.1080/10584609.2020.1840462>
- Chaiken, S. (1980). *Heuristic versus systematic information processing and the use of source versus message cues in persuasion*. 39(5), 752–766. <https://doi.org/10.1037/0022-3514.39.5.752>
- Chan, M. S., & Albarracín, D. (2023). A meta-analysis of correction effects in science-relevant misinformation. *Nature Human Behaviour*, 7(9), 1514–1525. <https://doi.org/10.1038/s41562-023-01623-8>

- Clemm Von Hohenberg, B., & Guess, A. M. (2023). When do sources persuade? The effect of source credibility on opinion change. *Journal of Experimental Political Science*, 10(3), 328–342. <https://doi.org/10.1017/XPS.2022.2>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Elsevier. <https://doi.org/10.1016/B978-0-12-179060-8.50006-2>
- Dias, N., Pennycook, G., & Rand, D. G. (2020). Emphasizing publishers does not effectively reduce susceptibility to misinformation on social media. *Harvard Kennedy School Misinformation Review*. <https://doi.org/10.37016/mr-2020-001>
- Ditto, P. H., Celniker, J. B., Siddiqi, S. S., Güngör, M., & Relihan, D. P. (2025). Partisan bias in political judgment. *Annual Review of Psychology*, 76(1), 717–740. <https://doi.org/10.1146/annurev-psych-030424-122723>
- Ecker, U. K. H., & Antonio, L. M. (2021). Can you believe it? An investigation into the impact of retraction source credibility on the continued influence effect. *Memory & Cognition*, 49(4), 631–644. <https://doi.org/10.3758/s13421-020-01129-y>
- Ecker, U. K. H., Lewandowsky, S., Cook, J., Schmid, P., Fazio, L. K., Brashier, N., Kendeou, P., Vraga, E. K., & Amazeen, M. A. (2022). The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1), 13–29. <https://doi.org/10.1038/s44159-021-00006-y>
- Finkel, E. J., Bail, C. A., Cikara, M., Ditto, P. H., Iyengar, S., Klar, S., Mason, L., McGrath, M. C., Nyhan, B., Rand, D. G., Skitka, L. J., Tucker, J. A., Van Bavel, J. J., Wang, C. S., & Druckman, J. N. (2020). Political sectarianism in America. *Science*, 370(6516), 533–536. <https://doi.org/10.1126/science.abe1715>

- Gawronski, B., Nahon, L. S., & Ng, N. L. (2025). Debunking three myths about misinformation. *Current Directions in Psychological Science*, 34(1), 36–42.
<https://doi.org/10.1177/09637214241280907>
- Gawronski, B., Ng, N. L., & Luke, D. M. (2023). Truth sensitivity and partisan bias in responses to misinformation. *Journal of Experimental Psychology: General*, 152(8), 2205–2236.
<https://doi.org/10.1037/xge0001381>
- Gawronski, B., Ng, N. L., Nahon, L. S., Hubeny, T. J., & Wurzinger, F. M. (2026). Susceptibility to misinformation as a signal detection problem. *European Review of Social Psychology*, 1–42. <https://doi.org/10.1080/10463283.2026.2675803>
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. Wiley.
- Hubeny, T. J., Nahon, L. S., Ng, N. L., & Gawronski, B. (2025). Who falls for misinformation and why? *Personality and Social Psychology Bulletin*.
<https://doi.org/https://doi.org/10.1177/01461672251328800>
- Institute for Quantitative Social Science. (2021, May 11). US media ownership. *The Future of Media Project*. <https://projects.iq.harvard.edu/futureofmedia/us-media-index>
- Kelley, H. H., & Michela, J. L. (1980). Attribution theory and research. *Annual Review of Psychology*, 31(1), 457–501. <https://doi.org/10.1146/annurev.ps.31.020180.002325>
- Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094–1096. <https://doi.org/10.1126/science.aao2998>

- Littrell, S., Meyers, E. A., & Fugelsang, J. A. (2024). Not all bullshit pondered is tossed: Reflection decreases receptivity to some types of misleading information but not others. *Applied Cognitive Psychology*, 38(1), e4154. <https://doi.org/10.1002/acp.4154>
- Macmillan, N. A., & Creelman, C. D. (2004). *Detection theory: A user's guide*. Taylor and Francis.
- Mang, V., Fennis, B. M., & Epstude, K. (2024). Source credibility effects in misinformation research: A review and primer. *Advances.in/Psychology*, 2(1), e443610. <https://doi.org/10.56296/aip00028>
- Modirrousta-Galian, A., & Higham, P. A. (2023). Gamified inoculation interventions do not improve discrimination between true and fake news: Reanalyzing existing research with receiver operating characteristic analysis. *Journal of Experimental Psychology: General*, 152(9), 2411–2437. <https://doi.org/10.1037/xge0001395>
- Mussweiler, T. (2003). Comparison processes in social judgment: Mechanisms and consequences. *Psychological Review*, 110(3), 472–489. <https://doi.org/10.1037/0033-295X.110.3.472>
- Nadarevic, L., Reber, R., Helmecke, A. J., & Köse, D. (2020). Perceived truth of statements and simulated social media postings: An experimental investigation of source credibility, repeated exposure, and presentation format. *Cognitive Research: Principles and Implications*, 5(1), 56. <https://doi.org/10.1186/s41235-020-00251-4>
- Nahon, L. S., Ng, N. L., & Gawronski, B. (2024). Susceptibility to misinformation about COVID-19 vaccines: A signal detection analysis. *Journal of Experimental Social Psychology*, 114, 104632. <https://doi.org/10.1016/j.jesp.2024.104632>

- Nedelcu, D., & Balaban, D. C. (2021). The role of source credibility and message credibility in fake news engagement. Perspectives from an experimental Study. *Journal of Media Research*, 14(3 (41)), 42–62. <https://doi.org/10.24193/jmr.41.3>
- Pehlivanoglu, D., Lin, T., Deceus, F., Heemskerk, A., Ebner, N. C., & Cahill, B. S. (2021). The role of analytical reasoning and source credibility on the evaluation of real and fake full-length news articles. *Cognitive Research: Principles and Implications*, 6(1), 24. <https://doi.org/10.1186/s41235-021-00292-3>
- Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592(7855), 590–595. <https://doi.org/10.1038/s41586-021-03344-2>
- Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188, 39–50. <https://doi.org/10.1016/j.cognition.2018.06.011>
- Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, 25(5), 388–402. <https://doi.org/10.1016/j.tics.2021.02.007>
- Peters, J. W. (2023, April 16). Fox News is on trial, and so are falsehoods about 2020. *The New York Times*. <https://www.nytimes.com/2023/04/16/business/media/fox-news-dominion-trial.html>
- Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. Springer New York.
- Petty, R. E., Wegener, D. T., & Fabrigar, L. R. (1997). Attitudes and attitude change. *Annual Review of Psychology*, 48, 609–647. <https://doi.org/https://doi.org/10.1146/annurev.psych.48.1.609>

- Pfänder, J., & Altay, S. (2025). Spotting false news and doubting true news: A systematic review and meta-analysis of news judgements. *Nature Human Behaviour*, 9(4), 688–699.
<https://doi.org/10.1038/s41562-024-02086-1>
- Pillai, R. M., & Fazio, L. K. (2021). The effects of repeating false and misleading information on belief. *WIREs Cognitive Science*, 12(6), e1573. <https://doi.org/10.1002/wcs.1573>
- Priester, J. R., & Petty, R. E. (1995). Source attributions and persuasion: Perceived honesty as a determinant of message scrutiny. *Personality and Social Psychology Bulletin*, 21(6), 637–654. <https://doi.org/https://doi.org/10.1177/0146167295216010>
- Roozenbeek, J., & van der Linden, S. (2019). The fake news game: Actively inoculating against the risk of misinformation. *Journal of Risk Research*, 22(5), 570–580.
<https://doi.org/10.1080/13669877.2018.1443491>
- Sanna, G. A., & Lagnado, D. (2025). Belief updating in the face of misinformation: The role of source reliability. *Cognition*, 258, 106090.
<https://doi.org/10.1016/j.cognition.2025.106090>
- Schaewitz, L., Kluck, J. P., Klösters, L., & Krämer, N. C. (2020). When is disinformation (in)credible? Experimental findings on message characteristics and individual differences. *Mass Communication and Society*, 23(4), 484–509.
<https://doi.org/10.1080/15205436.2020.1716983>
- Schwarz, N., & Jalbert, M. (2020). When (fake) news feels true: Intuitions of truth and the acceptance and correction of misinformation. In R. Greifeneder, M. E. Jaffé, E. J. Newman, & N. Schwarz (Eds.), *The psychology of fake news: Accepting, sharing, and correcting misinformation* (pp. 73–89). Routledge.

- Simchon, A., Zipori, T., Teitelbaum, L., Lewandowsky, S., & van der Linden, S. (2026). A signal detection theory meta-analysis of psychological inoculation against misinformation. *Current Opinion in Psychology*, 67, 102194. <https://doi.org/10.1016/j.copsyc.2025.102194>
- Sultan, M., Tump, A. N., Ehmann, N., Lorenz-Spreen, P., Hertwig, R., Gollwitzer, A., & Kurvers, R. H. J. M. (2024). Susceptibility to online misinformation: A systematic meta-analysis of demographic and psychological factors. *Proceedings of the National Academy of Sciences*, 121(47), e2409329121. <https://doi.org/10.1073/pnas.2409329121>
- Sultan, M., Tump, A. N., Geers, M., Lorenz-Spreen, P., Herzog, S. M., & Kurvers, R. H. J. M. (2022). Time pressure reduces misinformation discrimination ability but does not alter response bias. *Scientific Reports*, 12(1), 22416. <https://doi.org/10.1038/s41598-022-26209-8>
- Swire-Thompson, B., Kilgallen, K., Dobbs, M., Bodenger, J., Wihbey, J., & Johnson, S. (2024). Discrediting health disinformation sources: Advantages of highlighting low expertise. *Journal of Experimental Psychology: General*, 153(9), 2299–2313. <https://doi.org/10.1037/xge0001627>
- Taylor, S. E., & Fiske, S. T. (1978). Salience, attention, and attribution: Top of the head phenomena. *Advances in Experimental Social Psychology*, 11, 249–288. [https://doi.org/10.1016/S0065-2601\(08\)60009-X](https://doi.org/10.1016/S0065-2601(08)60009-X)
- Traberg, C. S., Harjani, T., Roozenbeek, J., & Van Der Linden, S. (2024). The persuasive effects of social cues and source effects on misinformation susceptibility. *Scientific Reports*, 14(1), 4205. <https://doi.org/10.1038/s41598-024-54030-y>

- Traberg, C. S., & van der Linden, S. (2022). Birds of a feather are persuaded together: Perceived source credibility mediates the effect of political bias on misinformation susceptibility. *Personality and Individual Differences, 185*, 111269.
<https://doi.org/10.1016/j.paid.2021.111269>
- Tversky, A., & Simonson, I. (1993). Context-dependent preferences. *Management Science, 39*(10), 1179–1189.
- Van Bavel, J. J., Harris, E. A., Pärnamets, P., Rathje, S., Doell, K. C., & Tucker, J. A. (2021). Political psychology in the digital (mis)information age: A model of news belief and sharing. *Social Issues and Policy Review, 15*(1), 84–113.
<https://doi.org/10.1111/sipr.12077>
- Van Bavel, J. J., & Pereira, A. (2018). The partisan brain: An identity-based model of political belief. *Trends in Cognitive Sciences, 22*(3), 213–224.
<https://doi.org/10.1016/j.tics.2018.01.004>
- Van Bavel, J. J., Rathje, S., Vlasceanu, M., & Pretus, C. (2024). Updating the identity-based model of belief: From false belief to the spread of misinformation. *Current Opinion in Psychology, 56*, 101787. <https://doi.org/10.1016/j.copsyc.2023.101787>
- van der Linden, S., Albarracín, D., Fazio, L., Freelon, D., Roozenbeek, J., Swire-Thompson, B., & Van Bavel, J. (2026). Using psychological science to understand and fight health misinformation: An APA consensus statement. *American Psychologist, 81*(5), 658–701.
<https://doi.org/https://doi.org/10.1037/amp0001598>
- Wallace, L. E., Wegener, D. T., & Petty, R. E. (2020). When sources honestly provide their biased opinion: Bias as a distinct source perception with independent effects on

credibility and persuasion. *Personality and Social Psychology Bulletin*, 46(3), 439–453.

<https://doi.org/10.1177/0146167219858654>

Zeng, H.-K., Lo, S.-Y., & Li, S.-C. S. (2024). Credibility of misinformation source moderates the effectiveness of corrective messages on social media. *Public Understanding of Science*, 33(5), 587–603. <https://doi.org/10.1177/09636625231215979>

Ziegler, R., Diehl, M., & Ruther, A. (2002). Multiple source characteristics and persuasion: Source Inconsistency as a determinant of message scrutiny. *Personality and Social Psychology Bulletin*, 28(4), 496–508. <https://doi.org/10.1177/0146167202287007>

Table 1

Potential effects of source credibility on judgments of true and false information in studies using incomplete designs that confound source credibility and information veracity (upper section) and studies using complete designs that do not confound source credibility and information veracity (lower section).

	Credible Source	Non-credible Source
Incomplete Design with Confound		
True Information	High hit rate	(missing)
False Information	(missing)	Low false-alarm rate
Complete Design without Confound		
True Information	High hit rate	Low hit rate
False Information	High false-alarm rate	Low false-alarm rate

Table 2

Participant demographics across experiments.

	Pilot	Expt. 1	Expt. 2a	Expt. 2b	Expt. 3	Expt. 4	Expt. 5	Expt. 6
<i>N</i>								
<i>n</i> Democrats	196	199	295	438	444	190	435	190
<i>n</i> Republicans	194	194	275	397	432	182	427	188
Gender								
Female	46.7%	43.8%	49.8%	49.8%	48.2%	55.4%	54.8%	56.9%
Male	52.3%	54.7%	49.1%	48.7%	50.5%	43.3%	43.6%	42.9%
Preferred not to answer	0.3%	0.5%	0.4%	0.2%	0.3%	0.5%	0.1%	0%
Other	0.8%	1.0%	0.7%	1.2%	1.0%	0.8%	1.5%	0.3%
Race								
White	88.2%	85.0%	85.1%	83.1%	83.6%	78.2%	72.2%	73.3%
Black or African American	5.1%	7.6%	6.3%	9.1%	10.6%	15.3%	15.1%	18.8%
Asian	6.9%	6.1%	9.8%	8.6%	5.9%	8.3%	13.2%	9.0%
American Indian or Alaska Native	2.6%	1.8%	1.1%	1.8%	1.0%	2.2%	1.6%	0.3%
Hawaiian or Pacific Islander	0.0%	0.0%	0.4%	0.2%	0.0%	0.5%	0.1%	0.3%
Other	2.6%	2.5%	1.9%	1.6%	2.1%	1.9%	1.7%	1.9%
Age								
Mean (SD)	42.7 (14.2)	39.9 (14.1)	39.5 (13.9)	37.6 (13.5)	41.4 (13.5)	42.9 (14.2)	40.6 (13.5)	42.6 (13.0)

Table 3

Means and standard deviations of participants' perceptions of familiarity, reliability, and political leaning for Forbes vs. National Enquirer and CNN vs. Fox News, Pilot Study.

Mean (<i>SD</i>)	Familiarity		Reliability		Political Leaning	
	Forbes	National Enquirer	Forbes	National Enquirer	Forbes	National Enquirer
Full Sample	4.22 (1.68)	4.01 (2.00)	5.11 (1.34)	2.74 (1.98)	4.14 (1.27)	4.12 (1.35)
Democrat	4.17 (1.53)	3.66 (1.94)	5.24 (1.28)	2.62 (2.01)	4.40 (1.14)	4.76 (1.18)
Republican	4.27 (1.82)	4.37 (2.01)	4.97 (1.38)	2.87 (1.94)	3.89 (1.34)	3.47 (1.20)
Mean (<i>SD</i>)	CNN		Fox News		Political Leaning	
	CNN	Fox News	CNN	Fox News	CNN	Fox News
Full Sample	5.76 (1.41)	5.74 (1.41)	4.26 (1.88)	3.52 (1.93)	2.34 (1.56)	6.29 (1.25)
Democrat	5.57 (1.39)	5.34 (1.50)	5.23 (1.17)	2.27 (1.44)	2.91 (1.54)	6.66 (0.93)
Republican	5.96 (1.41)	6.15 (1.18)	3.27 (1.96)	4.78 (1.49)	1.76 (1.35)	5.93 (1.41)

Note. Higher scores on perceived reliability indicate greater perceived reliability in reporting accurate information. Higher political leaning scores indicate pro-Republican leanings, and lower scores indicate pro-Democrat leanings.

Table 4

Sample sizes by political affiliation and means and standard errors of truth sensitivity across all experiments.

	Sample Size		Mean Truth Sensitivity (<i>SE</i>)		
	Democrat	Republican	Ideology- Congruent	Ideology- Incongruent	Apolitical
Experiment 1					
Source Absent	101	100	.84 (.05)	.75 (.05)	-
Source Present	98	94	.99 (.06)	.97 (.06)	-
Experiment 2a					
No Source	98	91	.84 (.05)	.79 (.05)	-
Reliable Source	98	88	.76 (.05)	.64 (.05)	-
Unreliable Source	99	96	.82 (.05)	.76 (.05)	-
Experiment 2b					
No Source	145	133	.71 (.04)	.69 (.04)	-
Reliable Source	145	134	.76 (.04)	.70 (.04)	-
Unreliable Source	148	130	.73 (.04)	.66 (.04)	-
Experiment 3					
No Source	146	146	.61 (.04)	.65 (.04)	1.00 (.04)
Reliable Source	150	142	.59 (.04)	.65 (.04)	.98 (.04)
Unreliable Source	148	144	.53 (.04)	.50 (.04)	.88 (.05)
Experiment 4					
Reliable Source	190	182	.65 (.04)	.74 (.04)	-
Unreliable Source	190	182	.64 (.04)	.69 (.04)	-
Experiment 5					
No Source	142	143	.71 (.04)	.70 (.04)	.83 (.04)
Congruent Source	149	141	.61 (.04)	.64 (.04)	.82 (.04)
Incongruent Source	144	143	.65 (.05)	.65 (.04)	.90 (.04)
Experiment 6					
Congruent Source	190	188	.55 (.04)	.61 (.04)	-
Incongruent Source	190	188	.58 (.04)	.60 (.04)	-

Table 5

Summary table detailing each experiment's preregistered hypotheses and results.

Experiment	SDT Index	Main Effect/ Interaction	Hypothesized Result	Actual Result	Hypothesis Supported?
1	Truth Sensitivity	Main Effect	$d'_{\text{source}} > d'_{\text{no source}}$	$d'_{\text{source present}} > d'_{\text{source absent}}$	Yes
	Partisan Bias	Interaction	$(c_{\text{ideology-incongruent}} - c_{\text{ideology-congruent}})_{\text{source}} < (c_{\text{ideology-incongruent}} - c_{\text{ideology-congruent}})_{\text{no source}}$	$(c_{\text{ideology-incongruent}} - c_{\text{ideology-congruent}})_{\text{source}} = (c_{\text{ideology-incongruent}} - c_{\text{ideology-congruent}})_{\text{no source}}$	No
2a	Truth Sensitivity	Main Effect	$d'_{\text{reliable source}} = d'_{\text{unreliable source}}$	$d'_{\text{reliable source}} = d'_{\text{unreliable source}}$	Yes
	Acceptance Threshold	Main Effect	$c_{\text{unreliable source}} > c_{\text{no source}} > c_{\text{reliable source}}$	$c_{\text{unreliable source}} = c_{\text{no source}} = c_{\text{reliable source}}$	No
	Partisan Bias	Main Effect	$c_{\text{ideology-incongruent}} > c_{\text{ideology-congruent}}$	$c_{\text{ideology-incongruent}} > c_{\text{ideology-congruent}}$	Yes
2b	Truth Sensitivity	Main Effect	$d'_{\text{reliable source}} = d'_{\text{unreliable source}}$	$d'_{\text{reliable source}} = d'_{\text{unreliable source}}$	Yes
	Acceptance Threshold	Main Effect	$c_{\text{unreliable source}} > c_{\text{no source}} > c_{\text{reliable source}}$	$c_{\text{unreliable source}} = c_{\text{no source}} = c_{\text{reliable source}}$	No
	Partisan Bias	Main Effect	$c_{\text{ideology-incongruent}} > c_{\text{ideology-congruent}}$	$c_{\text{ideology-incongruent}} > c_{\text{ideology-congruent}}$	Yes
3	Partisan Bias	Main Effect	$c_{\text{ideology-incongruent}} > c_{\text{ideology-congruent}}$	$c_{\text{ideology-incongruent}} > c_{\text{ideology-congruent}}$	Yes
	Acceptance Threshold	Interaction	Apolitical headlines: $c_{\text{unreliable source}} > c_{\text{reliable source}}$ Ideology-congruent headlines: $c_{\text{unreliable source}} = c_{\text{reliable source}}$	Across all types of headlines: $c_{\text{unreliable source}} = c_{\text{reliable source}}$	No

Ideology-incongruent headlines: $C_{\text{unreliable source}} = C_{\text{reliable source}}$					
4	Partisan Bias	Main Effect	$C_{\text{ideology-incongruent}} > C_{\text{ideology-congruent}}$	$C_{\text{ideology-incongruent}} > C_{\text{ideology-congruent}}$	Yes
	Acceptance Threshold	Main Effect	$C_{\text{unreliable source}} > C_{\text{reliable source}}$	$C_{\text{unreliable source}} > C_{\text{reliable source}}$	Yes
5	Partisan Bias	Preregistered	$C_{\text{ideology-incongruent}} > C_{\text{ideology-congruent}}$	$C_{\text{ideology-incongruent}} > C_{\text{ideology-congruent}}$	Yes
	Acceptance Threshold	Interaction	Ideology-congruent headlines: $C_{\text{congruent source}} > C_{\text{no source}} > C_{\text{incongruent source}}$ Ideology-incongruent headlines: $C_{\text{congruent source}} < C_{\text{no source}} < C_{\text{incongruent source}}$ Apolitical headlines: $C_{\text{congruent source}} = C_{\text{no source}} = C_{\text{incongruent source}}$	Across all types of headlines: $C_{\text{congruent source}} = C_{\text{incongruent source}} = C_{\text{no source}}$	No
6	Partisan Bias	Main Effect	$C_{\text{ideology-incongruent}} > C_{\text{ideology-congruent}}$	$C_{\text{ideology-incongruent}} > C_{\text{ideology-congruent}}$	Yes
	Acceptance Threshold	Interaction	Ideology-congruent headlines: $C_{\text{congruent source}} > C_{\text{incongruent source}}$ Ideology-incongruent headlines: $C_{\text{congruent source}} < C_{\text{incongruent source}}$	Across all types of headlines: $C_{\text{congruent source}} < C_{\text{incongruent source}}$	No

Figure 1. Graphical depiction of *truth sensitivity* within Signal Detection Theory, reflecting the distance between distributions of judgments about true and false information along the dimension of perceived veracity. Distributions that are closer together along the perceived-veracity dimension reflect a lower sensitivity, indicating that participants' ability in correctly discriminating between true and false information is relatively low (upper panel). Distributions that are further apart along the perceived-veracity dimension reflect a higher sensitivity, indicating that participants' ability in correctly discriminating between real news and fake news is relatively high (lower panel). Figure adapted from Gawronski, Nahon, and Ng (2025). Reprinted with permission.

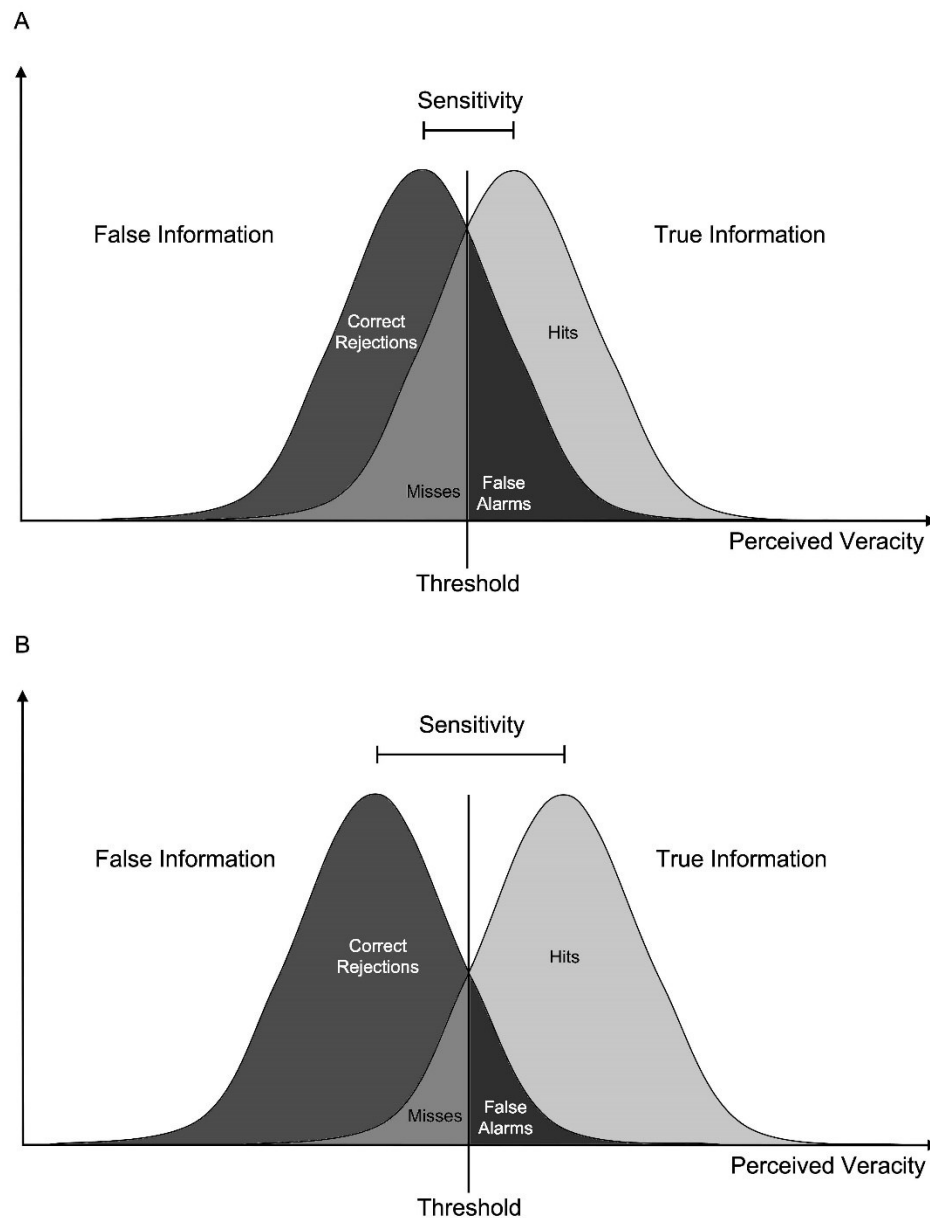


Figure 2. Graphical depiction of *acceptance threshold* within Signal Detection Theory, reflecting the threshold along the dimension of perceived veracity at which a participant decides to switch their decision. When judging information as true (vs. false), the threshold indicates the degree of veracity the participant must perceive before judging information as true. Any stimulus with greater perceived veracity than that value will be judged as true, whereas any stimulus with lower perceived veracity than that value will be judged as false. A low threshold would indicate that a participant is generally more likely to judge information as true (upper panel), whereas a high threshold would indicate that a participant is generally less likely to judge information as true (lower panel). Figure adapted from Gawronski, Nahon, and Ng (2025). Reprinted with permission.

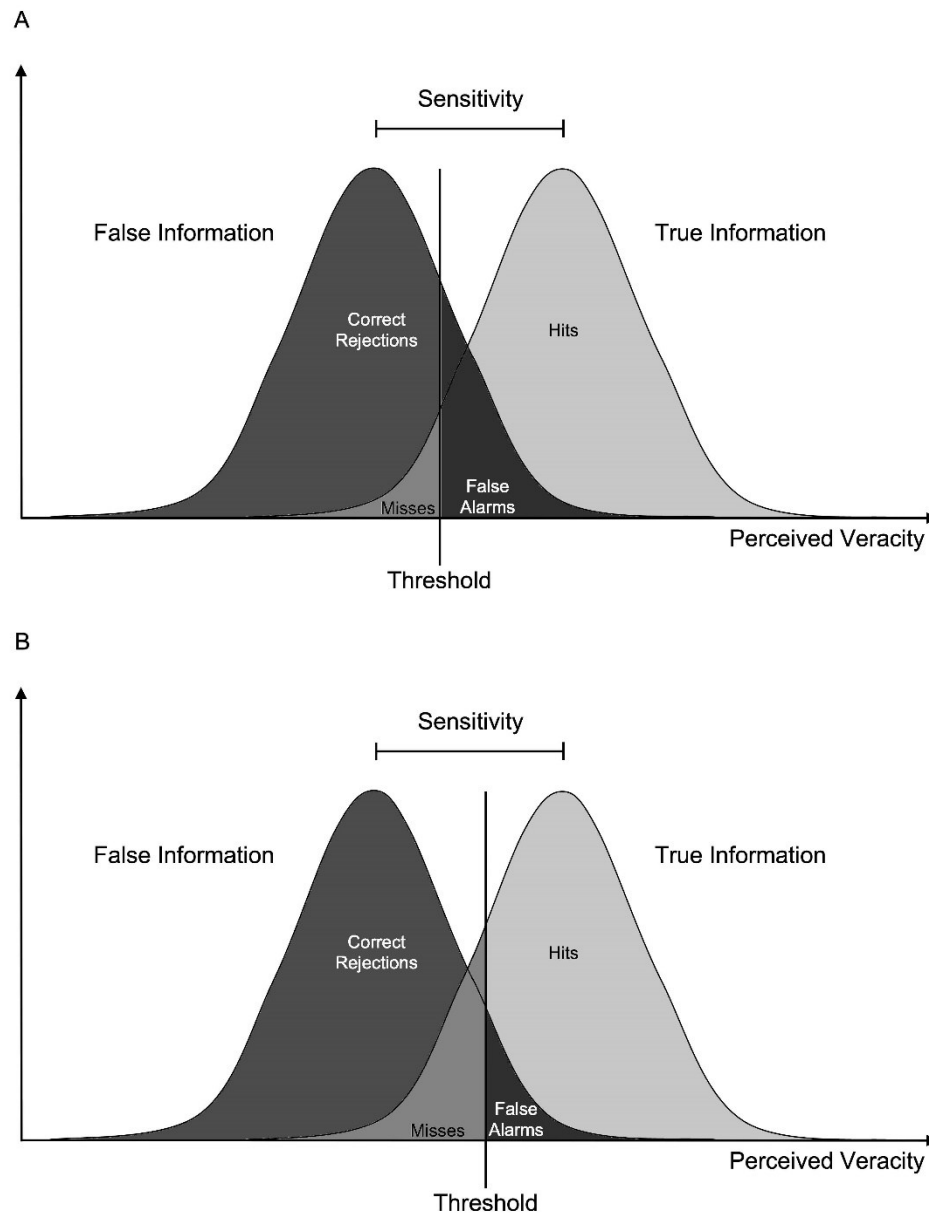
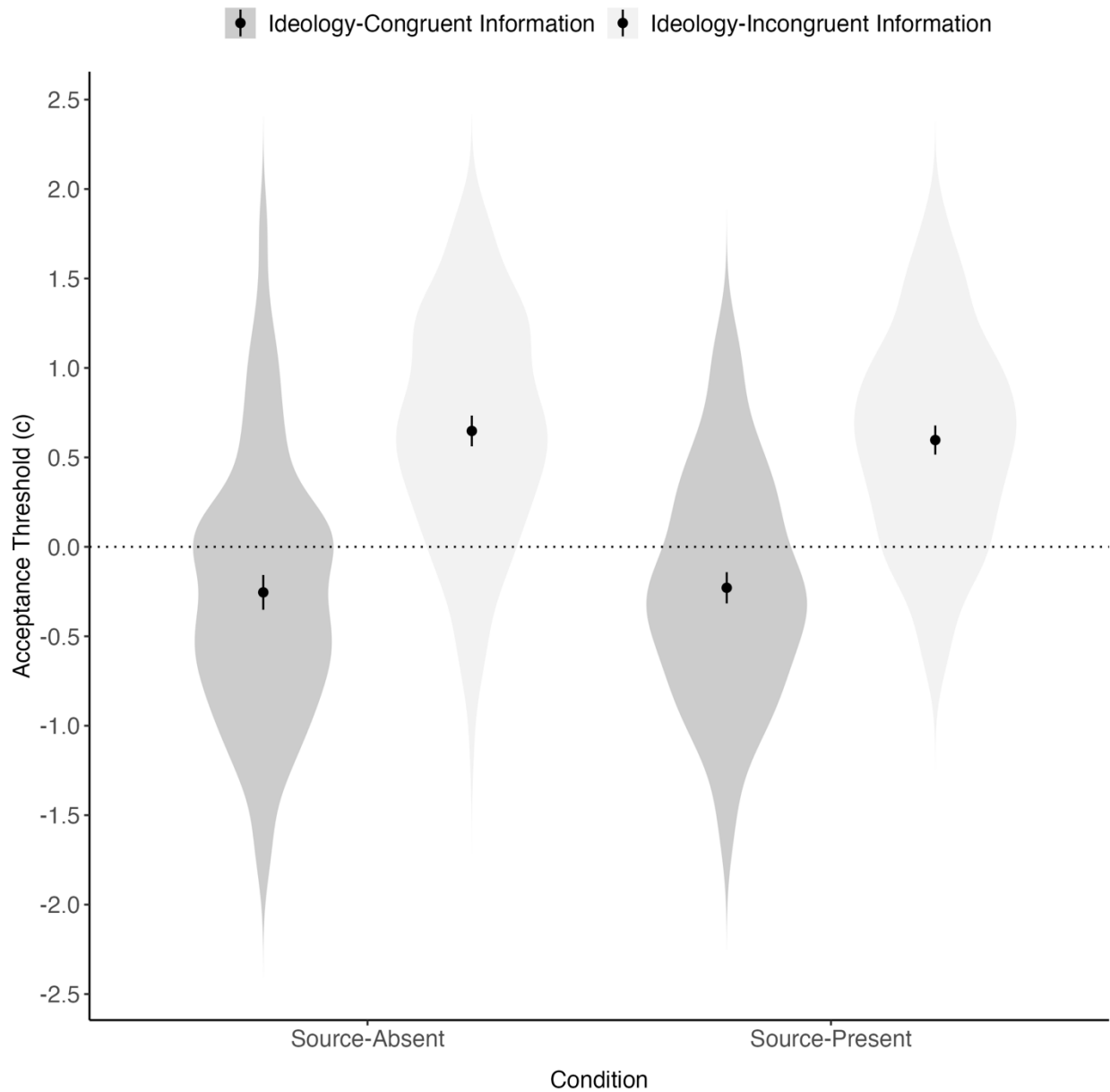
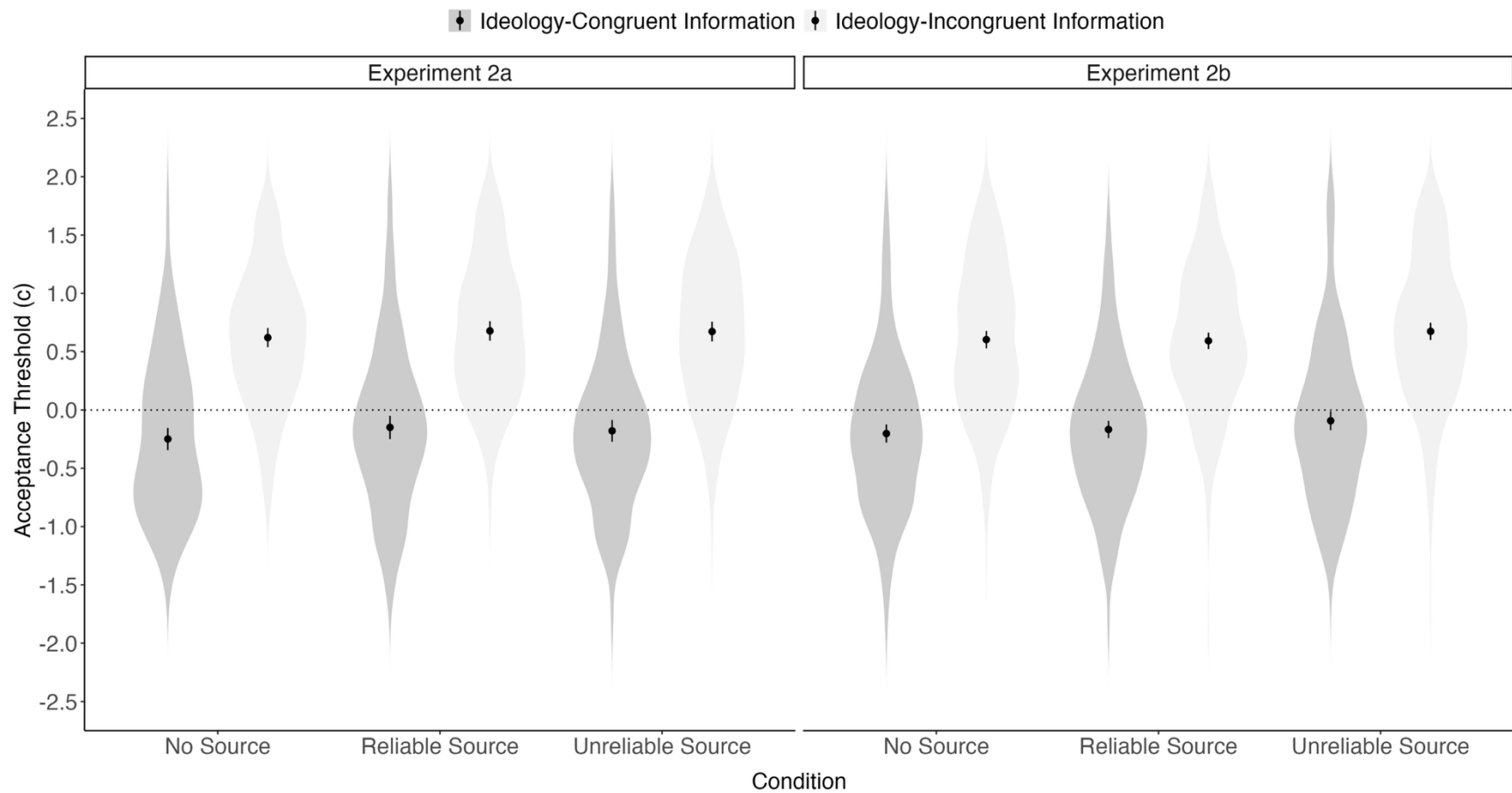


Figure 3. Signal detection c scores reflecting acceptance threshold in responses to political information as a function of headline ideology-congruence and source condition, Experiment 1.



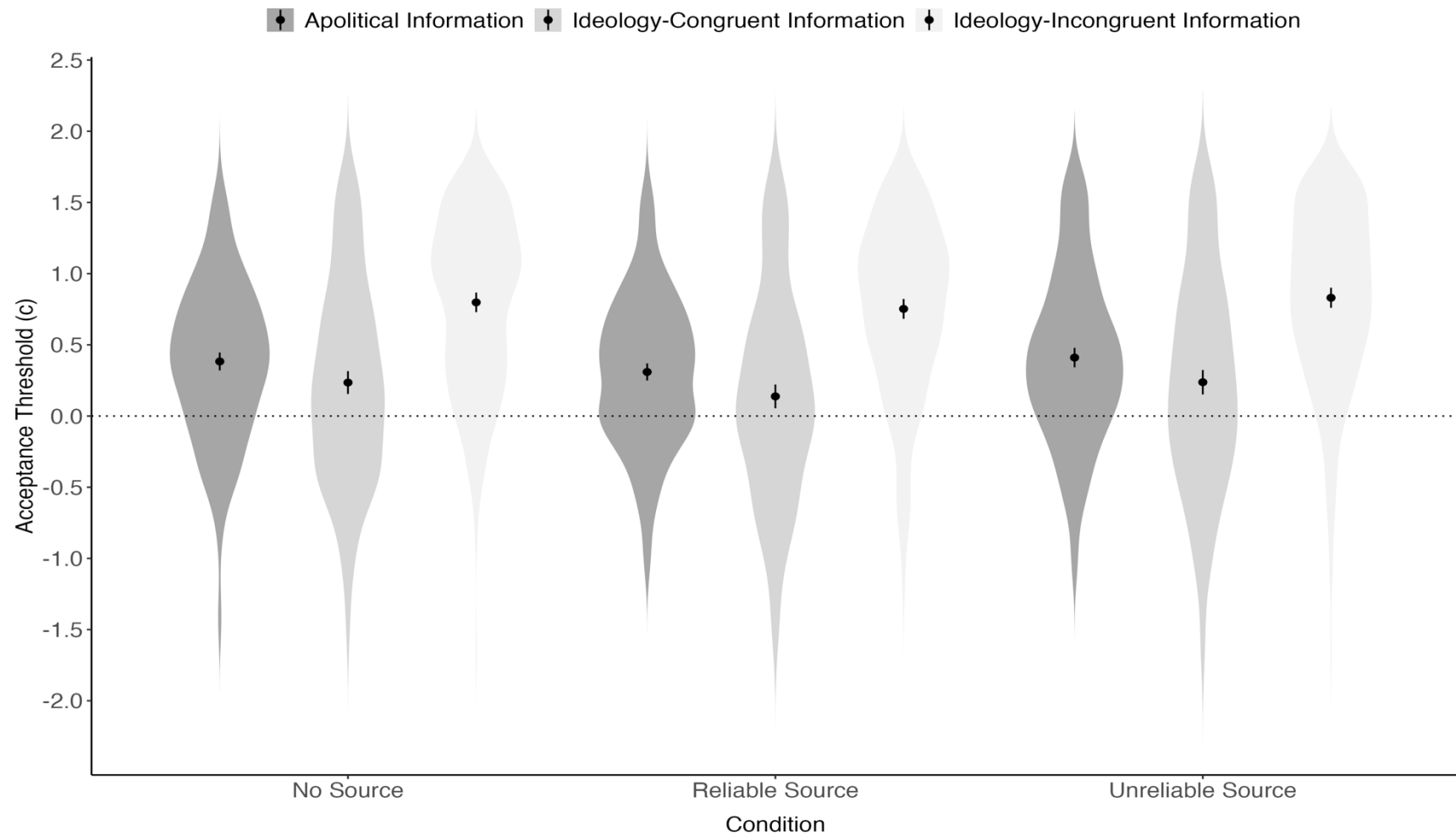
Note. Higher c scores reflect higher acceptance threshold. Error bars depict 95% confidence intervals.

Figure 4. Signal detection c scores reflecting acceptance threshold in responses to political information as a function of headline ideology-congruence and source condition, Experiments 2a and 2b.



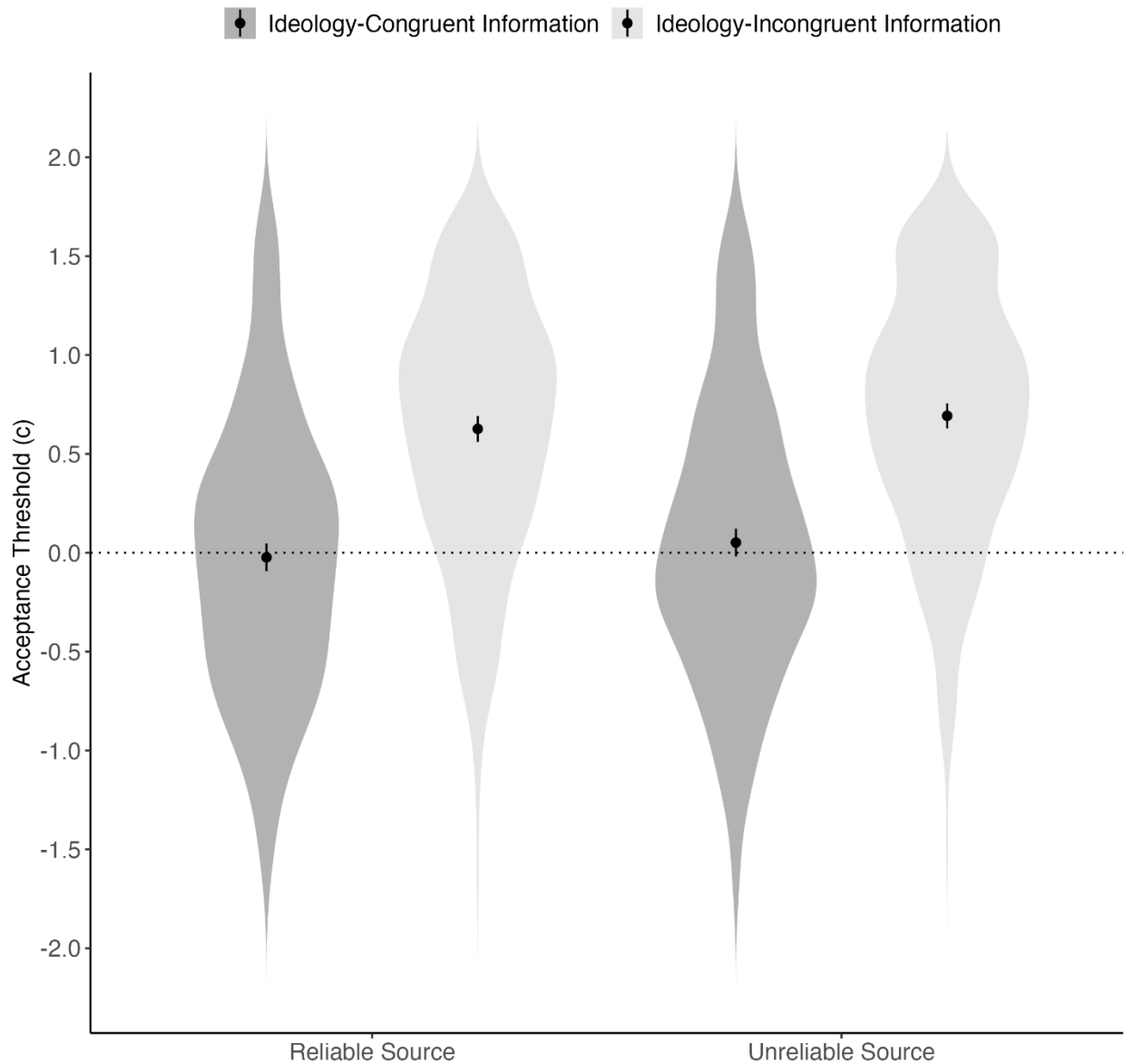
Note. Higher c scores reflect higher acceptance threshold. Error bars depict 95% confidence intervals.

Figure 5. Signal detection c scores reflecting acceptance threshold in responses to information as a function of headline slant and source condition, Experiment 3.



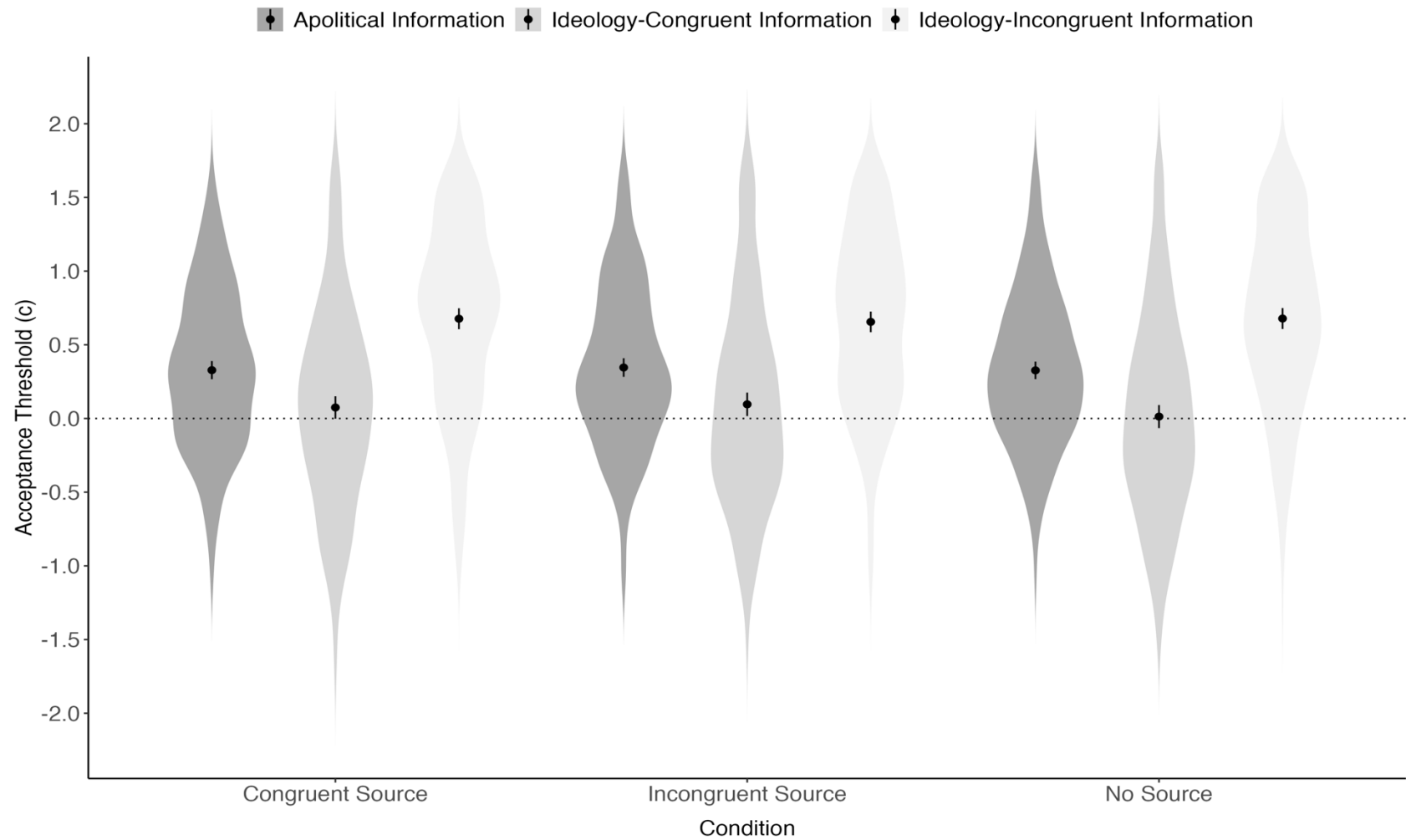
Note. Higher c scores reflect higher acceptance threshold. Error bars depict 95% confidence intervals.

Figure 6. Signal detection c scores reflecting acceptance threshold in responses to political information as a function of headline ideology-congruence and source reliability, Experiment 4.



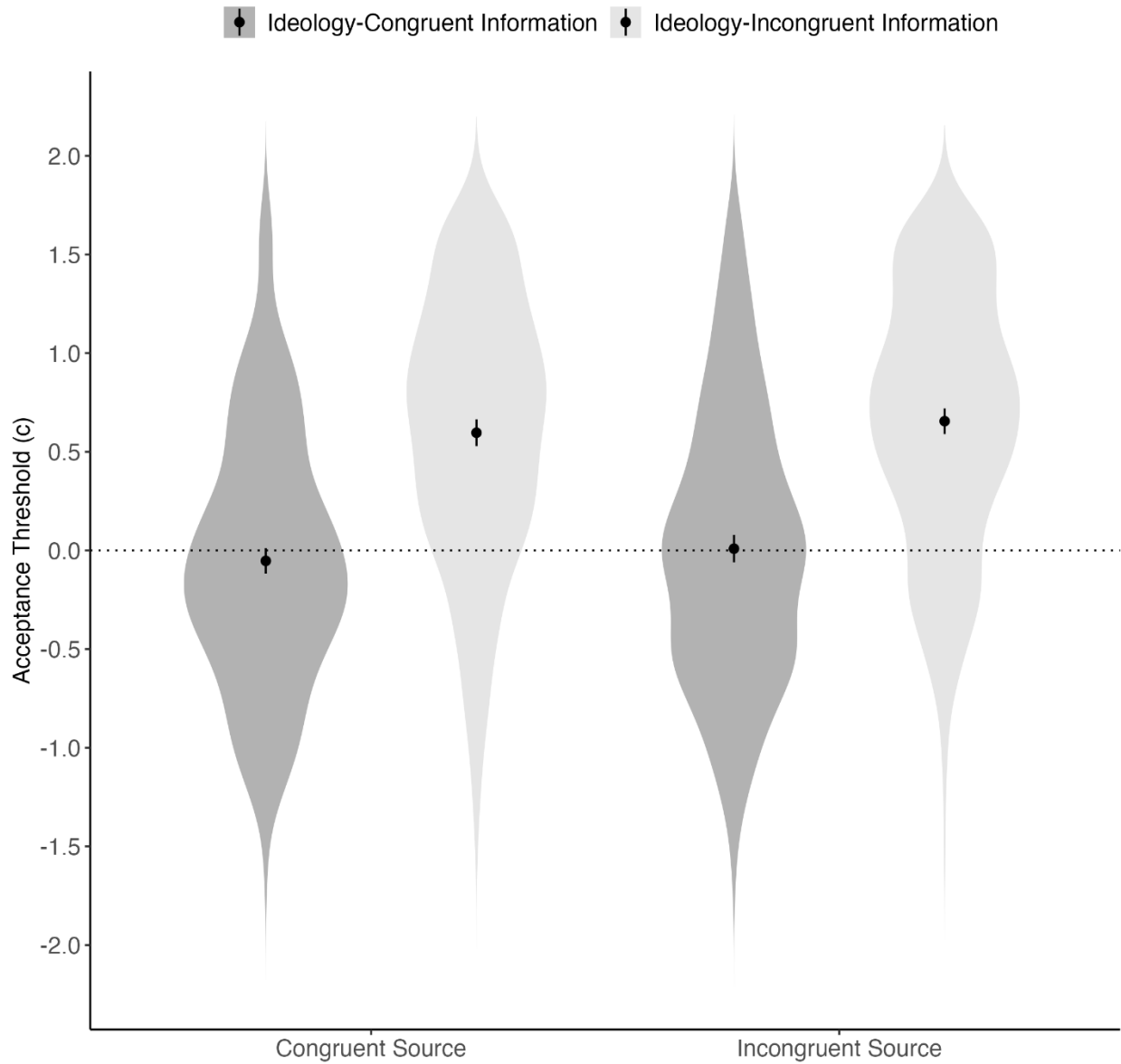
Note. Higher c scores reflect higher acceptance threshold. Error bars depict 95% confidence intervals.

Figure 7. Signal detection c scores reflecting acceptance threshold in responses to political information as a function of headline ideology-congruence and source partisanship, Experiment 5.



Note. Higher c scores reflect higher acceptance threshold. Error bars depict 95% confidence intervals.

Figure 8. Signal detection c scores reflecting acceptance threshold in responses to political information as a function of headline ideology-congruence and source partisanship, Experiment 6.



Note. Higher c scores reflect higher acceptance threshold. Error bars depict 95% confidence intervals.