UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

INSIGHTS INTO HOW STIMULUS ORDER, STIMULUS CLASS, AND INDIVIDUAL DIFFERENCES INFLUENCE PERFORMANCE FOR FACT-CHECKED TRUE AND FAKE NEWS HEADLINES

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

JUSTIN DURHAM

Norman, Oklahoma

2024

Downloaded from SHAREOK. Persistent Handle URL: https://hdl.handle.net/11244/340495

INSIGHTS INTO HOW STIMULUS ORDER, STIMULUS CLASS, AND INDIVIDUAL DIFFERENCES INFLUENCE PERFORMANCE FOR FACT-CHECKED TRUE AND FAKE NEWS HEADLINES

A DISSERTATION APPROVED FOR THE

DEPARTMENT OF PSYCHOLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Michael Wenger, Chair

Dr. Scott Gronlund

Dr. Hairong Song

Dr. Randa Shehab

© Copyright by JUSTIN DURHAM 2024

All Rights Reserved.

Acknowledgements

I owe the greatest thanks to my family – my wife, Nerida, and our 1-year-old daughter, Mia. To my wife, Nerida, your love, support, and understanding have been my rock while completing this journey. Your dedication for our family has provided me with the strength and stability needed to see this through. I could not have done this without you. To my 1-year-old daughter, Mia, you are my biggest accomplishment and greatest joy in life. Your laughter and natural curiosity have been a reminder of what truly matters. I also thank my parents and grandparents who have provided so much moral support and encouragement along the way. They have supported and helped me in more ways than I can express. I would like to thank my advisor, Dr. Scott Gronlund, for his guidance and mentorship throughout my doctoral training. Your expertise and insights have been instrumental in shaping this work. Finally, I wish to thank the members of my committee, Dr. Michael Wenger, Dr. Hairong Song, and Dr. Randa Shehab for their willingness to serve and provide valuable insights.

Table of Contents

Abstract vii
Background 1
Current Study 2
RQ1: How accurate are people at discriminating true from fake news headlines? 2
RQ2: Does stimulus order impact different aspects of performance? 4
RQ3: What individual differences predict confidence, discrimination, and response bias in detecting true news? 5
Overview 7
Methodology 7
Hypotheses 9
Discrimination-Focused. 9
Response Bias-Focused. 9
Confidence-Focused. 9
Method 10
Participants 10
Materials and Stimuli 10
Fake and True News Stimuli 10
Measures 12
Design 12
Procedure 12
Results 13
Outlier Detection 13
Descriptives 13
Factors 13
Individual Differences 16
Discrimination and Response Bias 16
Item level 16
Subject level 20
Confidence and Calibration 21
Individual Differences 22
Confidence Judgments 23
Discrimination 25
Bias 27
Discussion 29

Study 2 32
Overview 32
Method 33
Participants 33
Materials and Stimuli 33
Fake and True News Stimuli 33
Measures 34
Design 34
Procedure 35
Results 35
Outlier Detection 35
Descriptives 35
Factors 35
Individual Differences 38
Discrimination and Response Bias 38
Item level 38
Subject level 42
Confidence and Calibration 42
Individual Differences 43
Confidence Judgments 43
Discrimination 46
Bias 47
Discussion 49
General Discussion 51
Strengths and Limitations 57
Future Research 60
Conclusions 62
References 63
Appendix 70

Abstract

Objective: I evaluated 1) how well people can objectively discriminate fact-checked true and fake news headlines, 2) whether the order headline stimuli are presented to participants influences aspects of performance, and 3) what individual differences predict participants confidence judgments, discrimination, and response bias.

Background: There are multiple verbal theories explaining what makes people susceptible to misinformation. However, many of these studies focus on *perceived* accuracy rather than *objective* accuracy. Accuracy and discrimination should be influenced by other stimulus effects involving the order in which headlines are presented to participants and the degree to which headlines are true or fake.

Method: Stimuli used in this study consisted of fact-checked true and fake headlines obtained from Politifact.com – a fact-checking website. In study 1, the original headline source was removed and was included in study 2. In two studies, participants were first randomly assigned to stimuli order conditions (true-fake, fake-true, or random) and then rated sharing likelihood and confidence in judgments for 30 fact-checked true and fake headlines using a Likert-style scale. Finally, participants completed individual difference measures associated with misinformation susceptibility.

Results: I found more similarities than differences in my studies. In general, participants were able to discriminate true and fake headlines, but not very well, when subjected to signal detection analyses. Interestingly, I found that participants were better calibrated in their confidence judgments and accuracy for fake headlines than true headlines. I suspect this was driven by decreases in discrimination participants exhibited for headlines labeled as definitely true. Mixed effects models found that order only predicted participants' confidence judgments when the degree to which stimuli were true of fake was considered. Individual differences in conspiracy belief and political orientation consistently explained the variance in participants' performance when in competition with other measures. Moreover, individual differences in participants' cognition supported a classical cognitive account of misinformation susceptibility in study 1, while study 2 found support for an integrated motivational account.

Conclusion: Findings revealed that while participants struggled to discriminate true from fake headlines, they were better calibrated and accurate with fake headlines. The effect of stimulus order on confidence judgments was dependent on the stimulus class. Individual differences in conspiracy beliefs and political orientation consistently influenced performance. Methodological considerations for future research are discussed.

Keywords: misinformation, fake news, signal detection theory, individual differences, confidence, calibration, order, fact-checked, source, mixed effects modeling

Insights into How Stimulus Order, Stimulus class, and Individual Differences Influence Performance for Fact-Checked True and Fake Headlines

Propaganda, including fake news, has been around since the beginning of mass media communication. However, what has changed is that today, the majority of people consume their news and information through the internet, which has increased the spread of misinformation and susceptibility to disinformation. The spread of fake news, facilitated through social media, has the potential to distort public opinions and affect policy-making processes. In response to the rise of fake news on social media, national authorities have attempted to curb the spread of fake news using various regulatory techniques such as information correction, content removal or blocking, and criminal sanction (Helm & Nasu, 2021). Many of these techniques are difficult to implement and may be incompatible with freedom of speech. However, due to increased polarization and algorithm-driven content online, some assert that the harmful consequences of fake news may outweigh the benefits of free speech and some restriction on freedom of speech is inevitable to effectively combat fake news (see Helm & Nasu, 2021). While empirical research has been done to investigate the effects of misinformation on individuals' perceptions of true and fake news in the presence of various manipulations and inoculation methods, little work has been done to answer how accurate people are at discriminating between true and fake news using formalized methodologies. Addressing this gap will help better define the extent of the problem and the best path forward.

Background

What makes people susceptible to fake news? There are various accounts to explain this including partisanship, analytical ability, motivated reasoning, and prior exposure (see Batailler et al., 2022). Two distinct psychological accounts for what predicts misinformation susceptibility involve: 1) The classical reasoning or cognitive inattention account that assumes analytical ability (i.e., cognitive reflection test and numeracy) are the strongest predictors of misinformation susceptibility (Pennycook & Rand, 2020, 2021), and 2) the motivated reasoning account, which assumes the role of "myside bias" (i.e., open-minded thinking, Baron et al., 2022) and political partisanship (see Kahan et al., 2017) as the strongest predictors of misinformation susceptibility (Van Bavel et al., 2021; van der Linden, 2022). Research has found support for the cognitive account (Pennycook et al., 2022; Pennycook & Rand, 2019, 2021), as well as support for the motivational account (Roozenbeek et al., 2022). Moreover, research reanalyzing Pennycook and Rand's (2019) data has found that accurate identification increased as a function of cognitive reflection (particularly when true and fake news is congruent with one's political ideology) and politically incongruent headlines were judged as fake, despite their veracity, more often compared with politically congruent headlines (Batailler et al., 2022). These findings suggest that accounts of cognitive reflection, motivated reasoning, and partisan bias are not mutually exclusive.

Another important factor explaining why people are susceptible to misinformation, although more obvious and perhaps harder to control for, is how prior exposure increases the perceived accuracy of fake news, known as the illusory truth effect (Pennycook et al., 2018). A meta-analysis found that corrections to misinformation are less effective with repeated exposure (see Walter & Tukachinsky, 2020). Research has found that evidence-based decisions are driven more by prior beliefs than the quality of evidence (Michal et al., 2021). The effect of repeated exposure, combined with other accounts of misinformation susceptibility such as analytical thinking, can increase the coherence between prior beliefs and new information. This may lead to

inaccurate judgments (Bago et al., 2022) and can influence how people view fact-checking (Lazer et al., 2018). With this, prior exposure can influence identification of fake news by reducing the ability to accurately discriminate and may increase the tendency to judge previously seen news headlines as true regardless of veracity (Batailler et al., 2022).

In my study, the main focus is on examining how stimulus order impacts the capability to distinguish between true and fake news (discrimination), the willingness to endorse something as true or fake (response bias), and to understand how the direction in which stimuli are presented to participants impacts these aspects of performance. I also evaluate changes in participants' discrimination across fact-checked stimuli that vary in their magnitude of veracity or falsehood and examine how closely participants align their confidence judgments to the varied veracity of the stimuli (confidence/calibration). Finally, a secondary focus of my study is to identify which individual difference factors predict participants' discrimination, bias, and confidence judgments. My study will assess participants' intention to share news before asking about confidence. The decision to have participants make a sharing decision first is based on Epstein et al.'s (2021) findings that discrimination was more challenging when sharing was considered before accuracy. I chose to present stimuli to participants in a similar way because this is more representative of how social media platforms operate. Also, it keeps the focus on accuracy rather than social motivations.

Current Study

RQ1: How accurate are people at discriminating true from fake news headlines?

The current study focused on how well individuals perform in identifying the *objective* accuracy of stimuli. In other words, I examined how accurate an individual's perception of a headline is to actual fact-checked ratings. Previous research has primarily focused on describing participant's *perceived* accuracy (how accurate a headline is to an individual's subjective perception or confidence judgment). Studies that have focused on perceived accuracy have gleaned what makes people susceptible to believing in misinformation but has not gotten us closer to understanding how well people identify true and fake news. The current study examined how objectively accurate participants' perceptions of headlines that vary in their degree of truth and falsity were utilizing confidence judgments to understand discrimination and calibration (the relationship between confidence and accuracy) at a more granular level.

This study assessed objective accuracy with stimuli that are graded in their degree of truthfulness or falsehood rather than relying on stimuli from discrete classes (i.e., either true or false). The stimulus set I used is unique because it includes stimuli that have been labeled as definitely, likely, and probably, true or fake. It seems obvious that headlines vary in their degree of truthfulness or falsehood, but what is not clear is how well people are at detecting these gradations in true and fake headlines. By utilizing graded stimuli, I may find that people can discriminate definitely true from definitely fake news but are not very good at discriminating among stimuli that reflect lesser degrees of certainty (e.g., those that are probably true, likely true).

Research has shown that, when using a deception detection paradigm to evaluate whether participants can accurately discriminate true and fake news, participants exhibit poor discrimination, near chance, at distinguishing true and fake news headlines (average of 51% accuracy, see Luo et al., 2022). Luo et al. also found a deception bias (default perceptions towards news being fake most of the time) in their study 1 and a truth bias (perception of new

information being likely truthful) when considering whether greater Facebook likes increased accuracy in their study 2.

Because people believe in fake news (Pennycook & Rand, 2021), and fake news spreads farther and faster online than true news (Vosoughi et al., 2018), participants may mistakenly judge fake news as true at a greater rate than mistakenly judge true news as fake. However, it is possible to observe individuals mistakenly judge true news as fake at a greater rate. Put simply, while differences in errors may be explained by a truth bias, other research has found that errors can also be explained by a deception bias (Luo et al., 2022). Explanations for why these biases occur may be explained by individual differences, including prior exposure, which I will explore, and or other factors such as the use of news headline stimuli and the exclusion versus inclusion of source information that reflect the design of social media platforms. For example, results from this study may show that order impacts discrimination, which would have implications for how social media companies monitor and manage information that is posted or shared on their platforms.

RQ1A: I examined the relationship between participants' confidence judgments, discrimination, and response bias for true and fake headlines that varied in degree of truth and degree of falsity. I expected that participants would be able to discriminate true from fake news, but not very well (Batailler et al., 2022). I also expected discrimination and response bias to be impacted as a function of individual differences measures (see RQ3 for details). The relationship between confidence and accuracy has been investigated in other applied settings such as eyewitness identification (Gronlund & Benajmin, 2018). In eyewitness identification, a strong relationship exists between confidence and accuracy in some circumstances (for a review see Wixted & Wells, 2017). Given this, I might expect there to be a similarly strong relationship between confidence and accuracy regarding the participants' judgments about the graded stimuli I used. On the other hand, a strong relationship may not translate to the misinformation domain because incorrect information may, and often does, exist alongside correct information. Research has demonstrated that among a sample of German citizens, the accuracy of their confidence judgments in climate change knowledge was only around half of what it could be based on the accuracy of their knowledge (Fischer et al., 2019). These findings suggest that, at least in the context of climate change knowledge among citizens, the relationship between confidence and accuracy may be not well calibrated.

RQ1B: Individuals who have inaccurate or misinformed beliefs tend to be highly confident in their knowledge for a domain and reject counterevidence (Rapp & Withall, 2024). This can result in individuals being miscalibrated, wherein participants are more confident in their judgment than they are accurate. The role that confidence plays in misinformation susceptibility has only recently been investigated (see Rapp & Withall, 2024). Prior research has shown that confidence in one's beliefs can become better calibrated with feedback (Miller & Geraci, 2011). However, studying how misinformation influences confidence, and how to improve individual's calibration in their beliefs, is crucial for educational interventions given that belief in misinformation, particularly conspiracy theories, become stable in adolescence or even earlier (see Byrne et al., 2023). If participants in my study are calibrated, then they should have higher confidence for true headlines than fake and should report higher confidence in their decisions when judging headlines that are fact-checked with greater certainty (e.g., definitely true or fake) than lesser certainty (e.g., likely and probably true or fake).

RQ2: Does stimulus order impact different aspects of performance?

This study also focused on the effect of the order in which targets (true news headlines) and distractors (fake news headlines) are presented. News headlines are typically not shown to people in random order on social media platforms. Rather, the algorithmic-nature of these platforms often feed people more information similar to what they are already viewing, which means that when social media users are exposed to a fake news headline (either by search or through algorithm-driven content), they often are exposed to additional fake news headlines (which can increase belief in misinformation, Del Vicario et al., 2016; Lorenz-Spreen et al., 2020). This can be especially prevalent during a new information cycle for an event that is still unfolding (e.g., the early stages of the COVID-19 pandemic). For this reason, I examined how order impacts judgments of headlines and objective accuracy across trials. Specifically, I evaluated how aspects of performance (discrimination, bias, and confidence) change when fact-checked, graded true and fake headlines, are presented to participants in one of the following orders, from true to fake, fake to true, or in a random order.

Order can serve as a cue for how individuals self-organize their thoughts and make sense of information. Therefore, order may impact people's ability to discriminate true from fake news, but it is more likely to impact their willingness (i.e., response bias) to endorse a headline as true or fake (Wixted, 2020). Prior research in recognition memory examining between-subjects manipulations of stimulus or task difficulty (similar to order) found that participants had a more conservative response bias (higher criterion placement to make a true response) when making recognition judgments that included distractors that were semantically similar to targets (harder discriminations) versus participants who made recognition judgments that included semantically different distractors (Benjamin, 2005). Further, Benjamin and Bawa (2004), using a within-subjects design, found that participants adjust their criterion to be higher (more conservative) when completing an easy task followed by a difficult task, but do not adjust their criterion if the difficult task is completed first. Some research suggests that people do not adjust their criterion placement in response to these manipulations because it takes cognitive effort to adjust (Stretch & Wixted, 1998) or because adjustments in criterion lag changes in task difficulty (Brown & Steyvers, 2005). Meanwhile, other research has found that people can shift their criterion across trials under time pressure (Heit et al., 2003) and that people adopt a more liberal criterion placement for negatively-valenced stimuli than neutral or positively-valenced stimuli (because memory is more sensitive to negative stimuli) (Dougal & Rotello, 2007; see Rotello & Macmillan, 2007). The takeaway from this is that I expect that there are individual differences in tendencies to strategically shift criterion, but an individual's tendency appears to be a stable cognitive trait. That is, an individual's willingness to shift is a cognitive trait that is consistent within and across domains (Layher et al., 2020). In other words, some individuals can alter their decision strategy in response to target prevalence or consequences for errors, but others may be unwilling (or unable) to adjust their strategy to new evidence or the ordering of that evidence.

RQ2A: Based on previous research regarding the limited impact of task difficulty in recognition memory, order may be more likely to induce changes in response bias in the fake news domain. If the random order condition is more difficult than either the fake-true or true-fake orders, I expect participants to be more conservative to respond "true" to headlines in the random order condition. On the other hand, order effects might show that fake-true order induces a deception bias (most information is perceived as fake) resulting in a more liberal criterion to respond "fake", whereas a true-fake order might induce a truth bias (most information is

perceived as true), making participants more inclined to make a "true" response (see Luo et al., 2022).

RQ2B: Bias and discrimination generally are considered independent of one another, and it is possible that order does not impact participants response bias but does influence participants discrimination. If participants are not biased in how they make judgments across conditions, but differ in their discrimination, then discrimination might be impacted by how stimuli are ordered. For example, the fake-true condition could be more challenging than the true-fake and random condition resulting in worse discrimination. Conversely, I could find that a random order is more difficult than conditions in a specific direction. This prediction is supported by findings from Benjamin and Bawa (2005) and the assumption that order conditions introduce cognitive load wherein participants adjust their criterions to make judgments.

RQ2C: Interestingly, order also may impact the confidence participants assign to true and fake stimuli. If participants are induced with a truth bias in the true-fake order condition, this should result in higher mean confidence for true headlines relative to fake headline, while the opposite should be observed with a deception bias in the fake-true order condition – higher mean confidence for fake headlines relative to true headlines. Previous research shows that the truth bias is a strong phenomenon (Pantazi et al., 2018). Conversely, a deception bias may be prevalent; it has been found in previous research involving making judgments for social media news headlines (see Luo et al., 2022) by means of reducing trust in true news when individuals are exposed to greater rates of fake news (see Altay et al., 2023). Participants in the random order condition are likely to show confidence ratings somewhere in the middle due to the unpredictable pattern.

In sum, order could impact participants' confidence judgments, response bias and discrimination. If order condition influences response bias, the true-fake condition may induce a bias to respond true to stimuli. If the fake-true condition induces a deception bias, participants will have a greater tendency to respond fake to stimuli. The random condition may not induce a particular tendency but instead may result in participants being unbiased relative to other conditions. Order may also influence discrimination in that if initial exposure to highly accurate headlines leads to more accurate judgments, then participants in the true-fake condition should have higher discrimination than participants in the fake-true condition. Similar to response bias, I expect discrimination in the random condition to fall somewhere between the true-fake and fake-true conditions.

RQ3: What individual differences predict confidence, discrimination, and response bias in detecting true news?

This study also aims to investigate how discrimination and bias for identifying true news headlines is influenced by individual differences measures related to misinformation susceptibility, including: the cognitive reflection test (CRT, Frederick, 2005; Pennycook et al., 2020b), numeracy skills (Cokely et al., 2012; Roozenbeek et al., 2020; Schwartz et al., 1997), actively open-minded thinking (AOT, Baron 2019), information literacy test (ILT, Boh Podgornik et al., 2016; Jones-Jang et al., 2021), generic conspiracist belief scale (GCB-5, Brotherton et al., 2013; Kay & Slovic, 2023), political ideology, and social media use.

CRT is a common measure used as a proxy for reflexive analytical thinking (Frederick, 2005). Items on the CRT each have an incorrect intuitive response and require individuals to engage in reflection to obtain a correct answer (e.g., "A bat and a ball cost $1.10. The bat costs

$1.00 more than the ball. How much does the ball cost? ___ cents."). In alignment with the classical reasoning account of misinformation susceptibility, research has found that discrimination of true and fake news increases as a function of increasing scores on the CRT (Pennycook & Rand, 2019). Numeracy is also a factor involved in analytical thinking (Cokely et al., 2012; Schwartz et al., 1997) and has been found to be the strongest predictor (followed by trust in scientists) of lower susceptibility to COVID-19 misinformation across five countries (Roozenbeek et al., 2020). An example numeracy item includes, "Which represents the highest risk of something happening: 1 in 10, 1 in 100, or 1 in 1000" (Roozenbeek et al., 2020, 2022). Based on these findings, I expect CRT and numeracy scores to be associated with better discrimination of true from fake news.

The motivational account of misinformation susceptibility acknowledges that analytical ability serves a role in detection but emphasizes that myside bias and partisanship play a more central role and should be stronger predictors than CRT or numeracy. AOT, a measure sensitive to acceptance of myside bias (Baron 2019), is claimed to be one of the strongest predictors of misinformation susceptibility; political ideology as a measure of partisanship also should predict discrimination (see Van Bavel et al., 2021). An example AOT item includes participants' endorsements of statements like, "People should revise their conclusions in response to relevant new information." Roozenbeek et al. (2022) found that the AOT had a stronger correlation with discrimination than the CRT on the psychometrically validated misinformation susceptibility test (MIST, Maertens et al., 2022). In fact, AOT had the strongest correlation with discrimination, followed by political ideology, an individuals' confidence judgments, numeracy, and CRT.

These findings suggest that while AOT and CRT both measure reflective analytical ability, they assess distinct constructs of misinformation susceptibility, with AOT being superior (Roozenbeek et al., 2022). These results also are consistent with previous research that found a negative association between the AOT and conspiratorial beliefs, more pronounced among Democrats, while the CRT was weakly associated (Pennycook et al., 2020a). In my study, I will examine how discrimination and response bias change as a function of AOT, CRT, and other individual difference measures. Individuals who are higher in AOT, CRT, and numeracy should have better discrimination. Democrat participants are likely to have better discrimination than Republican participants, but political ideology should induce a stronger change on response bias than discrimination because stimuli that are ideologically-congruent can increase judgment ratings to respond true. Further, research has shown that Democrats report much higher levels of trust for news sources than Republicans (Pew Research Center, 2020, January 24).

I also examined how the generic conspiracist belief scale (GCB-5, Kay & Slovic, 2023) impacts discrimination and response bias. The GCB assesses conspiracy ideation, or the tendency to believe in conspiracy theories, and includes items such as, "The government permits or perpetrates acts of terrorism on its own soil, disguising its involvement." Research has found a relationship between political ideology and conspiratorial cognition (van der Linden, 2021) and found that participants with more conservative ideology, greater endorsement of conspiratorial beliefs, and poorer CRT scores, have greater belief in fake news headlines about election fraud (Calvillo et al., 2021). Given this, I expect people with high belief in conspiracist beliefs to report a conservative ideology and have worse discrimination.

Social media usage, a participant's quantifiable consumption of various social media platforms including Facebook, Instagram, Twitter/X and others, is also associated with belief in conspiracy theories and misinformation. Research has found that the relationship between social

media usage and belief in misinformation is contingent on conspiratorial cognition (Enders et al., 2021). Thus, it can be expected that greater social media use should be associated with conspiratorial beliefs and, therefore, worse discrimination. Participants were asked to rate how often they use each platform (13 in total) on a 1 *Never* to 5 *All the time* Likert scale. A mean score of usage across the 13 platforms included will represent participant's social media usage.

I also will examine how information literacy (ILT) impacts fake news identification. Concerns about the need for media literacy education have been raised (Clayton et al., 2020). The assumption behind the media literacy approach is that when people with greater media literacy consume false information, they do so in a more critical way, mitigating the influence of fake news (Jones-Jang et al., 2021). Previous research has found information literacy to be a significant predictor of identifying fake news stories whereas other forms of literacy (e.g., media, news, and digital literacy) were not (Jones-Jang et al., 2021). An example information literacy item includes, "The most reliable, verified, concise and comprehensive description of an unknown specialized concept can be found in _____." I expect discrimination to increase as a function of information literacy. It is also likely that greater information literacy should increase the amount of evidence people require to respond that a headline is true (i.e., increase response bias).

Overview

Participants were randomly assigned to an order condition, and I evaluated whether participants' discrimination and response bias are influenced by order. Participants were presented with a series of headlines and rated their likelihood to share a headline and then rated their confidence in how true or fake each headline is. Participants were instructed that there was a 50/50 chance that a news headline was true or fake on each trial. Participants also completed the aforementioned individual difference measures related to misinformation susceptibility. Two studies are conducted; both study 1 and study 2 are similar in concept and interested in understanding the relationship between participants' confidence judgments for news headline stimuli and their accuracy with graded, fact-checked ratings. The only difference between study 1 and study 2 is that the news source associated with each fact-checked stimuli are included in study 2 but not study 1.

Methodology

The methodology in the current research differs from previous studies investigating fake news in multiple ways. First, many studies investigating fake news record perceived accuracy as a binary response (yes or no). However, the veracity of information presented in news headlines – even those that are fact-checked – is not black and white. The current study examines perceived accuracy on an ordinal scale that allows participants to rate their confidence in a headline in a graded fashion. This affords the analysis of the relationship between participants' confidence judgments and objective accuracy (i.e., calibration). The accuracy of individual's confidence judgments in true and fake news identification is important to better understand because individuals who are better calibrated in their ability to assess veracity should be less susceptible to misinformation (Basol et al., 2020), and confidence has been found to have a reliable relationship with accurate identifications in other applied domains such as eyewitness memory (see Gronlund & Benjamin, 2018). Research has found that confidence judgments for true and fake news headlines do not meaningfully differ as a function of question framing (accuracy, trustworthiness, real or fake judgment, etc.), nor response mode (6-point, 7-point, and

binary scales) (Roozenbeek et al., 2022). This finding suggests that how confidence is assessed does not impact judgments of true and fake news. Therefore, in my study, I chose to have participants' rate how confident they are a headline is true or fake using a six-point ordinal scale.

Second, many stimuli used in previous studies have limitations, including being made up by the researchers, constrained to a specific period of political unrest (e.g., the 2016 US presidential election), only including text as stimuli, surrounding a single topic (e.g., COVID-19), and/or being exclusively binary (e.g., true or fake, right-leaning or left-leaning). These limitations raise issues regarding generalizing results to the social media environment and more fully understanding the processes underpinning what makes people susceptible to believing in fake news. The current study addresses these limitations by using representative stimuli from a fact-checking website (Politifact.com) that includes news headlines on a graded 'truthfulness' scale, a source, an associated image, and spans multiple topics across a range of years. Moreover, the stimuli are representative of what social media users see on platforms making the stimuli more realistic than making up my own stimuli or using computer-generated headlines.

Third, this study includes individual differences measures commonly used in previous research and explores which measures best predict discrimination and response bias in misinformation identification. Although some research has investigated how individual differences predict perceived accuracy judgments, only one I am aware of has predicted discrimination and response bias in conjunction with an individual difference measure (see Batailler et al., 2022). The current study uses multiple individual difference measures to predict discrimination and response bias across experimental manipulations.

I analyzed individuals' discrimination and bias using Signal Detection Theory (SDT; Macmillan & Creelman, 2004), a formalized model that disentangles discrimination (i.e., ability to distinguish between true and fake news headlines) and response bias (i.e., willingness to judge a given news headlines as true or fake) into separate parameters. In the context of evaluating true and fake news, response bias refers to the tendency of participants to favor one type of response over another, irrespective of the actual veracity of the stimuli. For instance, a conservative response bias means that participants set a high threshold for labeling something as true. This corresponds to a rightward placement of the criterion on a SDT graph. Conversely, a liberal response bias indicates that participants are more inclined to classify stimuli as true, setting a lower threshold for labeling something as true, which corresponds to a leftward placement of the criterion.

The signal detection framework is often used for evaluating discrimination and bias in experimental psychology and many other domains (e.g., eyewitness identification, medical decision making, aviation safety, human error; see Swets et al., 2000), but until recently has seen limited use by researchers investigating fake news (but see Barajas & John, 2023; Gawronski et al., 2023; Kemp et al., 2022; Modirrousta-Galian & Higham, 2023; Ross et al., 2018; Sultan et al., 2022). I also conducted receiver operating characteristic (ROC) analyses to assess an atheoretical measure of empirical discriminability (Wixted et al., 2017). Calibration analyses were performed to look at the relationship between participants' confidence judgments and their accuracy for true and fake news. Finally, I explored what predictors explain the variance in discrimination and response bias using mixed-effects models to control Type-I error rates.

Hypotheses

My expectations for study 1 are separated into three categories involving discrimination-based, response bias-based, and confidence-based hypotheses.

Discrimination-Focused.

- **Hypothesis 1**: Discrimination is expected to differ relative to stimulus class and confidence rating. For example, discrimination should be lower when comparing likely true and likely fake news headlines, but higher when comparing definitely true and definitely fake news headlines. The discrimination between definitely true or fake headlines and likely true or fake headlines should be larger than the relative discrimination between probably true and probably fake news headlines. Discrimination is also expected to be influenced by order. A fake-true order could prime participants to make judgments in a manner that results in lower discrimination relative to true-fake order.

  Hypothesis 2: Discrimination is expected to increase as a function of CRT, numeracy, AOT, and information literacy. Discrimination is expected to decrease as a function of conspiratorial cognition, social media use, and ideology (liberal – conservative scale). This hypothesis is consistent with previous research findings that these individual differences are associated with accuracy.

Response Bias-Focused.

- **Hypothesis 3**: Response bias or criterion placement is expected to differ by order. The true-fake condition is expected to induce a truth bias that should make participants set a lax criterion for responding "true" to headlines. This is because participants need less evidence to respond "true" when presented with all the true headlines prior to the fake headlines. The fake-true condition is expected to induce a deception bias that should make participants set a cautious criterion for responding "true" to headlines because being exposed to fake headlines prior to true headlines should induce conservative responding to true headlines. The random condition should be relatively unbiased because there is no pattern amongst the headlines by which to adjust one's criterion.
- **Hypothesis 4**: Participants high in CRT, numeracy, AOT, and information literacy should set a conservative criterion relative to participants low in these measures. Furthermore, participants high in conspiratorial cognition, social media use, and those that report a conservative ideology, should set a lax criterion relative to participants low in these measures and reporting a liberal ideology. This hypothesis is consistent with previous research findings that these individual differences are associated with how individual's respond to fake news with different tendencies.

Confidence-Focused.

- **Hypothesis 5**: Participants who are better calibrated in their confidence judgments should be less susceptible to misinformation in terms of their error rates. Overall, participants are expected to be more poorly calibrated when making confidence judgments for fake news headlines relative to true news headlines. This prediction is supported by previous research investigating the accuracy in citizens' climate change knowledge that found similar findings (see Fischer et al., 2019). It is also possible to find that participants are better calibrated for fake rather than true headlines. This prediction is supported by research finding participants have a deception bias for news on social media platforms

and a truth bias when stimuli credibility is manipulated with high or low counts of Facebook likes (Luo et al., 2022).

- **Hypothesis 6**: Confidence judgments should differ as a function of individual differences. Participants high in CRT, numeracy, AOT, and information literacy should rate fake stimuli with greater confidence on average relative to participants low in these measures. Furthermore, participants high in conspiratorial cognition, social media use, and reporting a conservative ideology should rate stimuli as true with greater confidence on average relative to participants low in these measures and those reporting a liberal ideology.

Method

Participants

In total, 524 participants enrolled in the online study through the University of Oklahoma's SONA system from 10/31/2023 to 12/11/2023. Data from 51 participants were removed for incomplete data resulting in a total of 473 completed cases for analysis. The majority of participants were between the ages of 18-24 (n = 465, 98%), female (n = 375, 80%), and White/Caucasian (n = 294, 62%). Participants most frequently reported being moderately (n = 176, 37%) or slightly (n = 163, 34%) familiar/knowledgeable about local, national, and world news. Participants most frequently reported their political orientation as unsure or neither conservative nor liberal (n = 120, 25%) followed by slightly conservative (n = 97, 20%), moderately liberal (n = 81, 17%), moderately conservative (n = 80, 17%), slightly liberal (n = 68, 14%), extremely conservative (n = 19, 4%), and extremely liberal (n = 8, 2%). Participants most frequently reported being slightly familiar/knowledgeable (n = 212, 45%) with the news stimuli that were presented in the study, followed by moderately familiar/knowledgeable (n = 121, 26%), and a percentage of participants had no knowledge/familiarity at all (n = 69, 15%).

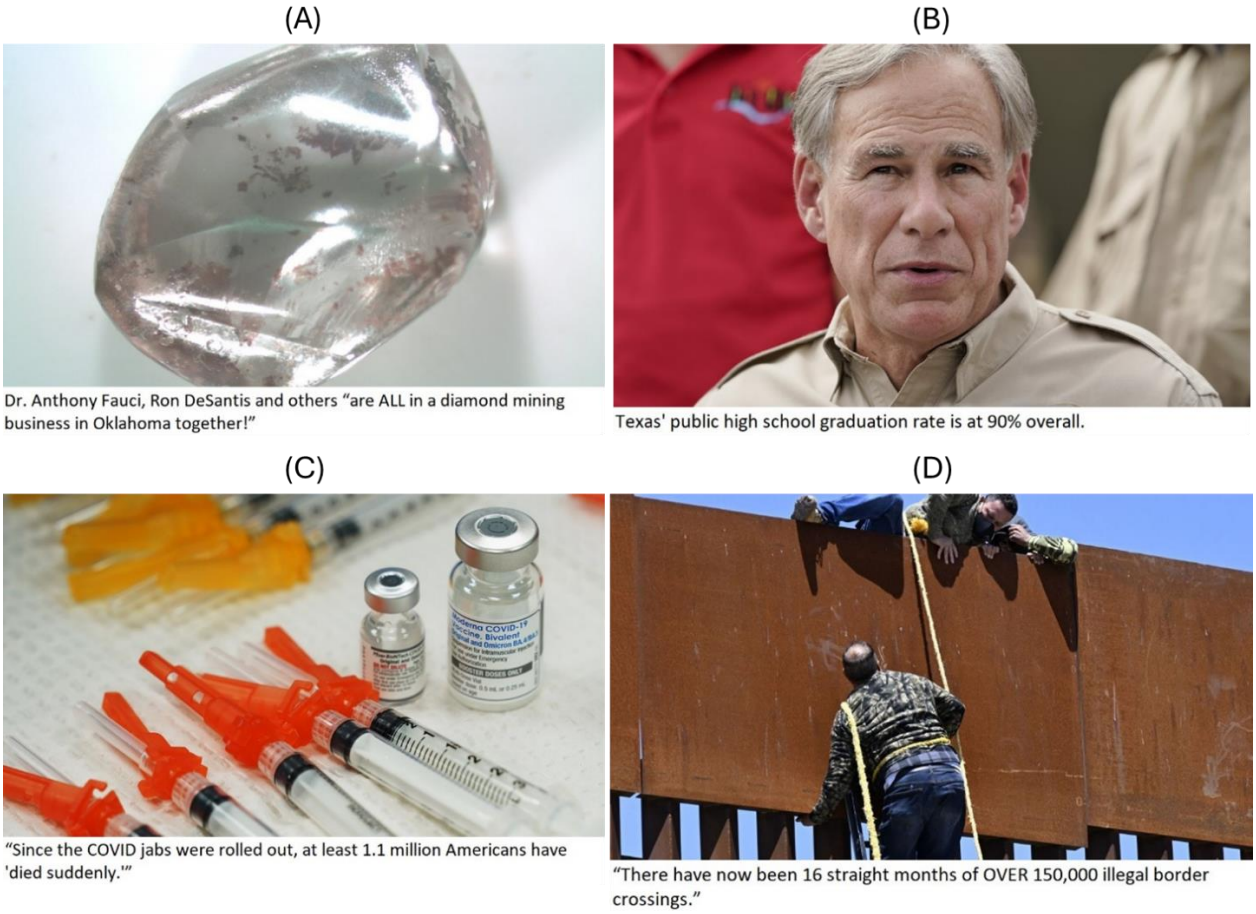
Materials and Stimuli

Fake and True News Stimuli

Fake and true news stimuli were identified and adapted from Politifact.com – a fact-checking website that rates the accuracy of claims made by elected officials and other sources on its 'truth-o-meter'. Politifact.com's truth-o-meter includes true and fake statements ranging from True, Mostly True, Half True, Mostly False, False, to Pants on Fire. Politifact.com headlines used in this study are related to economic, social, and political topics. Stimuli included headline text, a static image relating to the headline text, and date of the headline (see Figure 1). Fifteen headlines from each of the six truth-o-meter categories were selected resulting in a total of 90 headline stimuli (45 true news headline stimuli and 45 fake news headline stimuli). The fact-checking labels from Politifact.com were changed to a Likert scale that represented their confidence in how true or fake a headline was. Stimuli were posted online on dates ranging from 06/21/2019 to 02/03/2023.

Figure 1

Examples of Fact-Checked Headline Stimuli



Note. (A) is an example stimuli of a fact-checked news headline labeled as definitely fake, (B) is an example stimuli of a fact-checked news headline labeled as definitely true, (C) is an example stimuli of a fact-checked news headline labeled as likely fake, and (D) is an example stimuli of a fact-checked news headline labeled as likely true.

A norming study with 20 volunteers was conducted in which they rated how confident the 90 headline stimuli were either fake or true. The top five stimuli from each stimulus class with the highest count of correct responses to their respective stimulus class (e.g., responding definitely true to a definitely true stimuli) were selected, resulting in a total of 30 fake and true news stimuli used. These results were also corroborated using ChatGPT 3.5 shortly prior to data collection. The text from each stimulus and the scale to rate each stimulus was included in the prompt to ChatGPT. ChatGPT was prompted to make judgments about the extent to which stimuli were true or fake using the same confidence scale participants used. The ChatGPT output was relatively consistent with results from the norming study in that the stimuli ChatGPT rated correctly in a respective stimulus class (e.g., responding definitely true to a definitely true stimuli) were very similar to those stimuli that participants rated the most frequently in the same stimulus class. Consequently, no adjustments to the stimuli assignments were made.

Measures

Participants completed the 3-item numeric (Frederick, 2005) and the 4-item non-numeric (Thomson & Oppenheimer, 2016) CRT. A composite score of the 7-items was used as participants' overall cognitive analytical ability. A participant's numeracy skills was measured with a composite score from three different numeracy tests: The 3-item Schwartz test (Schwartz et al., 1997), the 4-item pencil and paper Berlin numeracy test (Cokely et al., 2012), and responses to the following item: "What represents the highest chance of something happening: 1 in 10, 1 in 1000, or 1 in 100" (Roozenbeek et al., 2020, 2022). Participants also completed the 10-item actively open-minded thinking scale (AOT, see Baron et al., 2022) using a 1-5 ordinal agreement scale, the 5-item version of the generic conspiracy belief scale (GCB-5, Kay & Slovic, 2023) using a 1-5 ordinal agreement scale, 5-items from the information literacy test (ILT, see Jones-Jang et al., 2021) using a multiple-choice format. A composite score was used for the AOT, GCB-5, and ILT. Measures used in these studies have been used to examine belief in misinformation and detection (Pennycook & Rand, 2019). A complete list of the measures' items is included in the Appendix.

Design

The study design was a 6 stimulus type (Definitely Fake, Likely Fake, Probably Fake, Probably True, Likely True, and Definitely True) x 3 stimulus order (True-Fake, Fake-True, or Random). Stimulus type was a within-subjects factor and stimulus order was a between-subjects factor. Participants were randomly assigned to an order condition. Participants confidence ratings (1: Definitely Fake–6: Definitely True) was the primary outcome or dependent variable in this study.

Procedure

Prior to beginning the study online, participants completed an informed consent form and then were provided with the following study instructions:

> *During the study, you will rate how confident you are news headlines are fake or true and how likely you would be to share it online. There is an equal likelihood that any headline is fake or true. The aim is to complete the trials as quickly and accurately as possible. You will complete practice trials before completing test trials. Finally, you will complete a series of questionnaires and survey items.*

First, participants completed the eight questions from the MIST-8 (Maertens et al., 2022) in a random order as practice trials. Next, participants were given the following instructions for the experimental trials:

> *You will do the following for each headline:*
>
> *1. Rate how likely you would be to share the post on a social media platform and*
>
> *2. Rate how confident you are the headline is fake or true.*

In the experiment, participants viewed the 30 stimuli in the randomly assigned order (True-Fake, Fake-True, or Random). As mentioned, participants made two responses for each stimulus: How likely they would be to share the post on social media and how confident they were the headline was either fake or true. Participants rated their likelihood to share on a 1-6 Likert scale (Not at all likely-Extremely likely) and rated their confidence on a 1-6 Likert scale (Definitely Fake-Definitely True)

After completing responses to all 30 stimuli, participants responded to the following item using a 1-6 Likert scale (Not at all familiar/knowledgeable-Extremely familiar/knowledgeable), "Rate how familiar you were with any of the previously seen news headlines presented." Participants also rated their familiarity/knowledge on a total of 10 distinct news-related domains (e.g., Climate Change and Environmental Sustainability, Current National Events, Economy/Business and Finance, Social Issues/Human Rights, etc.) using the same 1-6 Likert scale. Next, participants completed the CRT and numeracy skill items in a random order followed by the GCBS, ILT, and AOT in a random order. Finally, participants rated how often they used various social media platforms (e.g., Facebook, TikTok, Instagram, Twitter/X, YouTube) on a 1-5 Likert scale (Never-All the time) and demographic items including age, race/ethnicity, political orientation, familiarity/knowledge with news (local, national, and world), education level, and gender. Finally, participants were presented with a debriefing page that informed them that they were done with the study and would receive course credit.

Results

Outlier Detection

Data were screened for outliers and adjusted in two ways:

1. Participants that had only used one side of the scale (responded true or fake for all stimuli) were removed. This resulted in the removal of data from 3 participants.
2. Boxplots of discrimination ability (*d'*; described below) identified 3 outliers that were removed.

With these six outliers identified and removed, there were a total of 467 cases remaining for analysis.

I considered removing participants with negative discrimination ability, which indicated that some participants were sensitive to the stimuli but in a way that produced incorrect responses. However, I chose not to because the current study is not a recognition memory study and negative *d'* values do not signal 'below chance' discrimination. Rather, negative *d'* values indicate that some stimuli are confusable enough that participants made numerous errors when they, ostensibly, should not have. Moreover, analyzing the data without negative *d'* participants did not change the global results described in these results; a larger proportion of participants had very strong discrimination relative to those with negative discrimination, and a previous study observed negative *d'* when reanalyzing data from multiple studies (see Batailler et al., 2022).

Descriptives

Factors

The average confidence rating, proportion correct, and average sharing rating grouped by condition, stimulus type, and stimulus class are reported below.

When the data were grouped by order condition, an ANOVA found significant group mean differences in participants' confidence judgments, $F(2, 464) = 3.34$, $p = 0.04$, $\eta^2 = 0.01$. Post hoc comparisons using Tukey's HSD test found significant differences between the T-F ($M = 3.62$, $SD = 1.37$) and F-T ($M = 3.49$, $SD = 1.29$) conditions, and between the T-F and Random ($M = 3.52$, $SD = 1.36$) conditions, but no significant differences between the Random and F-T conditions. This result suggests that participants in the T-F order, on average, had significantly

higher confidence judgments relative to participants in the F-T order condition. An ANOVA test found significant differences between the proportion of correct responding across conditions, $F(2, 464) = 7.9$, $p < 0.05$, $\eta^2 = 0.03$. Post hoc comparisons using Tukey's HSD tests found that the proportion of correct responding was significantly lower in the F-T ($M = 0.60$, $SD = 0.49$) condition relative to both the T-F ($M = 0.64$, $SD = 0.48$) and Random ($M = 0.63$, $SD = 0.48$) condition.

When the data were grouped by the stimulus type, a paired sample *t*-test found that participants had significantly different mean confidence judgments, $t(466) = 25.26$, $p < 0.05$, $d = 1.18$, with confidence for Fake ($M = 3.93$, $SD = 1.35$) stimuli being rated statistically higher, on average, compared to confidence for True ($M = 3.16$, $SD = 1.21$) stimuli. A paired sample *t*-test found that participants had significantly different rates of correct responding, $t(466) = 3.91$, $p < 0.05$, $d = 0.46$, $d = 0.18$, with the proportion of correct responding being higher for True ($M = 0.65$, $SD = 0.48$) stimuli compared to Fake ($M = 0.59$, $SD = 0.49$) stimuli on average.

When the data were grouped by the fact-checked stimuli class, an ANOVA found significant group mean differences in participants' confidence judgments, $F(5, 2796) = 189.1$, $p < 0.05$, $\eta^2 = 0.25$. Post hoc comparisons using Tukey's HSD test found significant differences between all comparisons except when comparing Probably Fake ($M = 3.84$, $SD = 1.37$) and Likely Fake ($M = 3.82$, $SD = 1.3$) stimuli. An ANOVA test found significant differences between participants' rate of correct responding, $F(5, 2796) = 28.48$, $p < 0.05$, $\eta^2 = 0.05$. Post hoc comparisons using Tukey's HSD test found significant differences in correct responding for all comparisons except probably true and definitely fake, likely true and definitely fake, definitely true and probably fake, and likely true and probably true. Means and standard deviations for each outcome by factor can be found in Table 1.

Table 1

Sample Size (n), Average Confidence (SD), p(Correct) (SD), and Average Sharing (SD) By Condition, Stimulus Type, and Stimulus Class

Condition	*n*	Average Confidence (SD)	p(Correct) (SD)	Average Sharing[1] (SD)
F-T	158	3.49 (1.29)	0.60 (0.49)	1.63 (1.14)
Random	155	3.52* (1.36)	0.63* (0.48)	1.88* (1.34)
T-F	154	3.62* (1.37)	0.64* (0.48)	1.75 (1.20)

Note. The asterisk indicates that the value was significantly different from the F-T condition.

Stimulus Type	*n*	Average Confidence (SD)	p(Correct) (SD)	Average Sharing[2] (SD)
True	467	3.16 (1.21)	0.65 (0.48)	1.86 (1.31)
Fake	467	3.93* (1.35)	0.59* (0.49)	1.64* (1.15)

Note. The asterisk indicates that the value was significantly different from True stimuli.

	Stimulus Class	*n*	Average Confidence (SD)	p(Correct) (SD)	Average Sharing[3] (SD)
1	Definitely True	467	3.34 (1.24)	0.59 (0.49)	1.82 (1.25)
2	Likely True	467	2.97* (1.09)	0.73* (0.44)	1.87 (1.34)
3	Probably True	467	3.17* (1.26)	0.64* (0.48)	1.89 (1.34)
4	Probably Fake	467	3.84* (1.37)	0.57 (0.50)	1.65* (1.13)
5	Likely Fake	467	3.82* (1.30)	0.57 (0.50)	1.67* (1.18)
6	Definitely Fake	467	4.13* (1.37)	0.65* (0.48)	1.61* (1.13)

Note. The asterisk indicates that the value was significantly different from Definitely True stimuli.

[1] An ANOVA test found significant differences between participants' sharing judgments across conditions, $F(2, 464) = 3.19$, $p = 0.04$, $\eta 2 = 0.01$. Post hoc comparisons using Tukey's HSD tests found significant differences between the Random ($M = 1.88$, $SD = 1.34$) and F-T ($M = 1.63$, $SD = 1.14$) conditions, but no significant differences between the T-F and other conditions.

[2] A paired sample *t*-test found that participants had significantly different mean sharing judgments, $t(466) = 9.88$, $p < 0.05$, $d = 0.46$, with likelihood to share being higher for True ($M = 1.86$, $SD = 1.31$) stimuli compared to Fake ($M = 1.64$, $SD = 1.15$) stimuli.

[3] An ANOVA test found significant differences between participants' likelihood to share across the stimuli classes, $F(5, 2796) = 7.74$, $p < 0.05$, $\eta 2 = 0.01$. Post hoc comparisons using Tukey's HSD test found significant differences in likelihood to share definitely true and definitely fake, likely true and definitely fake, probably true and definitely fake, likely true and likely fake, probably true and likely fake, probably fake and likely true, and probably true and probably fake stimuli.

Individual Differences

Descriptive statistics for the individual difference measures are shown in Table 2. Accuracy scores (presented in mean column) on the CRT, Numeracy skills, and ILT measures, showed that overall, participants were around chance accuracy (e.g., 50/50). Higher average scores on the GCB-5 and AOT indicate higher overall ratings of truthfulness for conspiracy theories (closer to definitely true) and greater agreement with active open-mindedness items. The sample did not have high average scores of social media usage with the platforms assessed in this study ranging from rarely-sometimes usage. On average, participants rated being unsure or neither liberal nor conservative political orientations (n = 117), followed by slightly, moderately, or extremely conservative (n = 194) and slightly, moderately, or extremely liberal (n = 156). Most participants were slightly to moderately familiar with and knowledgeable about local, national, and world news. Estimates of skewness and kurtosis did not indicate substantial nonnormality.

Table 2

Sample Size (n), Mean, Standard Deviation (SD), and Median by Individual Difference Measure

Measure	*n*	Mean	SD	Median
Cognitive Reflection Test	467	0.43	0.26	0.43
Numeracy Skills	467	0.56	0.24	0.5
Information Literacy Test	467	0.41	0.22	0.4
General Conspiracist Beliefs	467	2.83	0.73	3
Active Open-Minded Thinking	467	3.74	0.45	3.7
Media Use Composite	467	2.27	0.41	2.23
Political Orientation	467	4.12	1.49	4
News Knowledge	467	2.78	0.96	3

Discrimination and Response Bias

Signal detection analyses were performed using participants' confidence judgments for stimuli. Participants' discrimination and bias estimates are reported at the item level and subject level, overall, and by order condition. Response rates were adjusted for perfect or zero discrimination. For response rates that had perfect discrimination, estimates were adjusted using t-1. Response rates of zero were adjusted using 1/t (where t =30 is total count of stimuli trials) (see MacMillan & Creelman, 2004; Stanislaw & Todorov, 1999).

Item level

How well could participants discriminate between the different stimulus classes? Table 3 shows the probability of responding fake or true across all participants. Two different estimates of *d'* are reported in Table 3; *d' adjacent* estimates the discrimination between one stimulus class (i.e., probably fake) and an adjacent stimulus class (i.e., probably true); *d' cumulative* describes the cumulative discrimination ability between the reference class (i.e. definitely fake) and

another stimulus class (likely true). Interestingly, a negative *d'* adjacent was observed for definitely true stimuli relative to likely true stimuli. This indicates that participants were sensitive to the stimuli but in a way that resulted in them making more errors, which is reflected in the lower overall cumulative discrimination between definitely fake and definitely true stimuli (*d'* = 0.60). If we removed all responses for definitely true stimuli, our cumulative *d'* across all the items would be ~1, suggesting participants can reasonably discriminate amongst the news headline stimuli.

The probability of responding true (i.e., p(True)) across the stimulus classes is plotted in Figure 2 to illustrate this deficiency when participants responded to headline stimuli fact-checked as definitely true. The expectation is that the probability of responding true should increase linearly across stimulus class. However, there is a large decrease in the proportion of making a true response when the stimuli are fact-checked as being highly truthful.

Thus far, my results have summarized the variables included in my study. The subsequent analyses are conducted to evaluate my hypotheses. Specifically, I test whether discrimination for fact-checked headlines that are labeled as definitely true and definitely fake are larger than those that are labeled as likely or probably true and fake (Hypothesis 1), and whether response bias differs by order (Hypothesis 3) in the Discrimination and Response Bias section. In addition, I evaluate how calibrated participants are in their confidence judgments about true and fake stimuli (Hypothesis 5) in the Confidence and Calibration section. Finally, I explore what factors may explain individuals' discrimination (Hypothesis 2), response bias (Hypothesis 4), and confidence judgments (Hypothesis 6) in the Individual Differences section.

Table 3

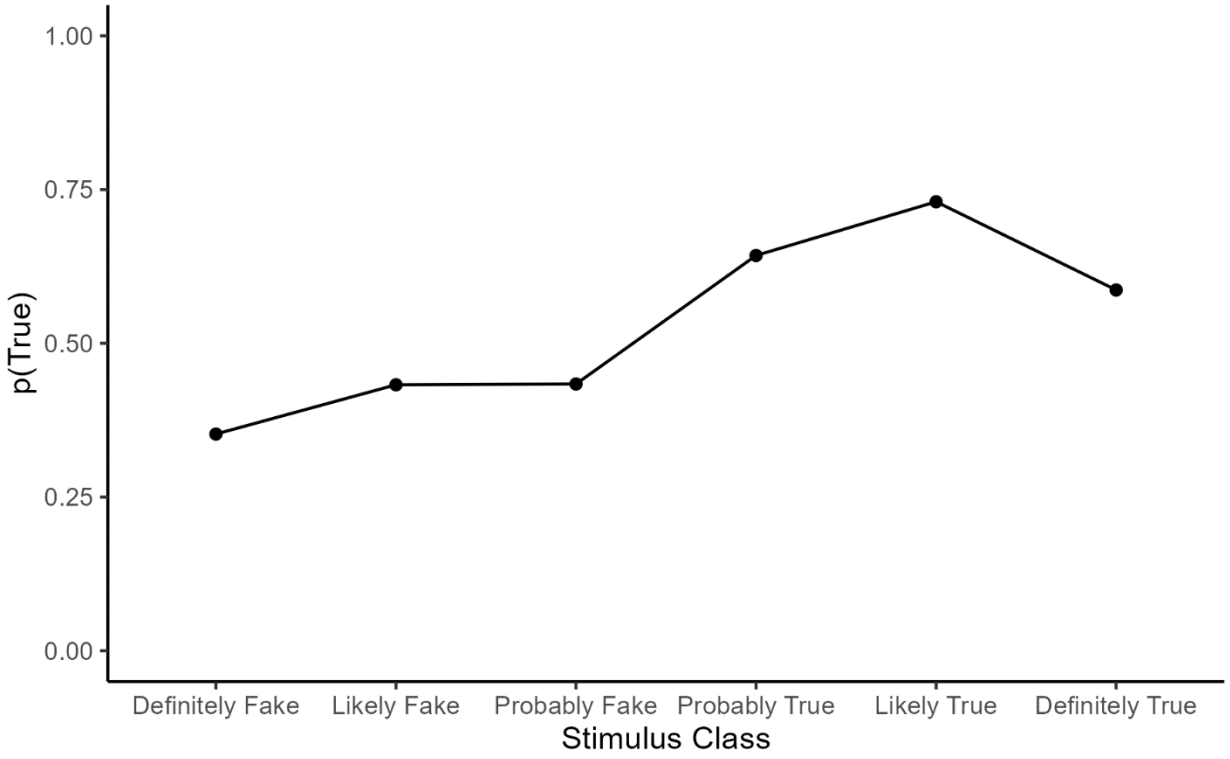
Probability of Responding Fake and True, d' Adjacent, and d' Cumulative by Stimulus Class

	Stimulus Class	p(Fake)	p(True)	d' Adjacent	d' Cumulative
1	Definitely Fake	0.648	0.352	NA	NA
2	Likely Fake	0.567	0.433	0.209	0.209
3	Probably Fake	0.566	0.434	0.003	0.212
4	Probably True	0.357	0.643	0.533	0.745
5	Likely True	0.270	0.730	0.247	0.992
6	Definitely True	0.413	0.587	-0.394	0.598

Note. $n = 467$.

Figure 2

Probability of Responding True Across Stimulus Class



I assessed accuracy and error rates as a function of stimulus type to address whether participants show lower discrimination comparing likely true and fake headlines and higher discrimination comparing definitely true and fake headlines (Hypothesis 1). Results in Table 4 suggest that discrimination is lower for stimuli fact-checked as probably fake or true (d' = 0.53) relative to the other stimulus classes and types (likely and definitely fake and true). Discrimination for probably fake and true stimuli was significantly lower than discrimination for likely fake and true stimuli, $t(466) = 15.01$, $p < 0.05$, and discrimination for definitely fake and true stimuli $t(466) = 2.87$, $p < 0.05$. However, inconsistent with my expectation, discrimination for likely fake and true stimuli was significantly higher compared to definitely fake and true stimuli (d' = 0.78), $t(466) = 11.11$, $p < 0.05$; and participants' discrimination for definitely fake and true (d' = 0.60) was only slightly higher than discrimination for the probably fake and true stimuli. This finding provides minor support for my first hypothesis that discrimination should increase with certainty in the fact-checked classes (probably, likely, and definitely), but suggests that something about the headlines fact-checked as definitely true stimuli made discrimination more challenging for participants, which resulted in a lower discrimination between the definitely fake and true stimuli pairs.

Table 4

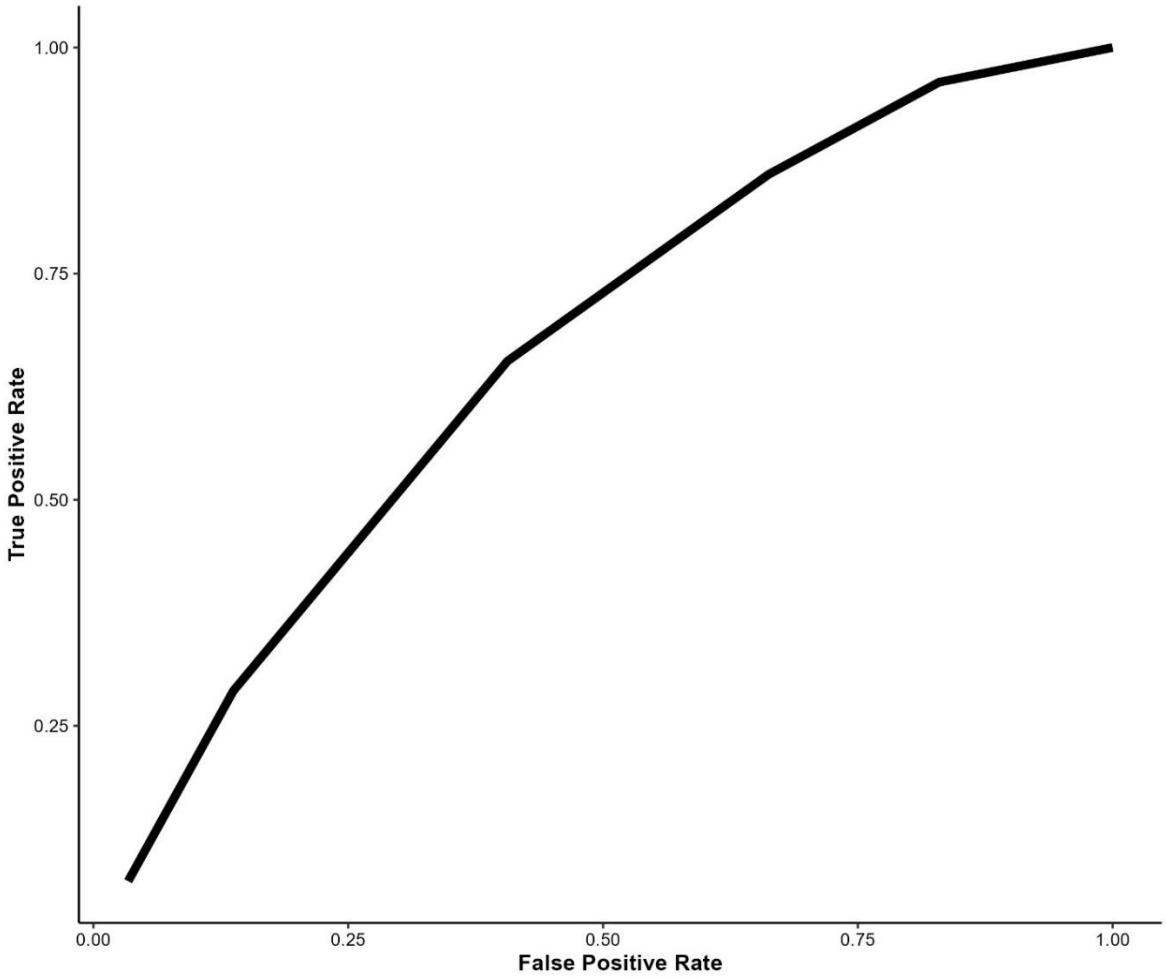
Discrimination (d') Estimates by Stimulus Class Pairs for Definitely, Likely, and Probably Fake and True Stimuli

Stimulus Class	*d'*
Definitely Fake vs Definitely True	0.598
Likely Fake vs Likely True	0.783
Probably Fake vs Probably True	0.533

Receiver operator characteristic (ROC) curves provide another measure for assessing individuals' ability to discriminate between fact-checked true and fake headlines. ROC analysis is important because it can be used to extract "an atheoretical measure of empirical discriminability" (Wixted et al., 2017). ROC curves plot "the hit rate versus the false alarm rate across different levels of response bias, holding discriminability constant" (Wixted, 2020). In my study, a "hit" or true positive, is defined as making a true response when presented with true stimuli (whether probably, likely, or definitely). A "false alarm" or false positive, is defined as making a true response when presented with fake stimuli (whether probably, likely, or definitely). The ROC curve plots the cumulative conditional probabilities of the hit rate on the Y-axis and false alarm rates on the X-axis at across different levels of confidence. The more bowing towards the top left corner of the space the curve has, the greater the discriminability between true and fake headlines. From the ROC curve in Figure 3, the area under the curve (AUC) represents a scale-invariant measure of discrimination across the ROC criterion thresholds. An AUC of 0.60 was estimated, which was significantly above chance, $t(466) = 4.28$, $p < 0.05$. These results indicate that individuals can categorize the stimuli at above chance levels, but are, nevertheless, relatively poor classifiers for the fact-checked true and fake headlines used in this study. These findings indicate that, overall, the test was challenging, and participants had a hard time discriminating between the true and fake news headlines.

Figure 3

ROC Curve for All Participants



Subject level

Table 5 shows the response types, response rates, and signal detection measures (*d'* and *c,* see Macmillan & Creelman, 2004; Stanislaw & Todorov, 1999) for the overall sample ($n = 467$) and separated by the order conditions. Overall, participants could discriminate between true and fake news, but not very well ($d' = 0.63$), and they were marginally biased ($c = -0.08$) to respond fake to stimuli.

I also predicted that discrimination would differ by order (Hypothesis 1). An ANOVA test did find significant group mean differences in discrimination across condition, $F(2, 464) = 6.89$, $p < 0.05$, $\eta^2 = 0.03$. Post hoc analyses showed that discrimination was significantly lower for the F-T condition compared to the T-F condition, $t(464) = 3.3$, $p < 0.05$ and Random condition, $t(464) = 3.11$, $p < 0.05$. This suggests that the order in which participants are presented with the fact-checked stimuli influences their ability to discriminate between true and fake news.

An ANOVA was performed to address whether response bias differed by order (Hypothesis 3). I predicted that the true-fake condition would induce a truth bias, the fake-true condition would induce a deception bias, and the random condition would be unbiased. However, I found no significant group mean differences in response bias across order. Participants presented with stimuli in a T-F order had the highest discrimination and were relatively unbiased in their responding. Participants presented with stimuli in an order from fake to true (F-T) had the lowest discrimination and had a slight tendency to respond Fake to stimuli. Participants presented with stimuli in a Random order had discrimination similar to the T-F condition but also had bias similar to those in the F-T condition.

Table 5

Counts of Participants Response Types, Response Rates, Sensitivity (d'), and Response Bias (c)

	n	Hits	FAs	CRs	Misses	HR	FAR	*d'* [95% CI]	*c* [95% CI]
Overall	467	4576	2846	4159	2429	0.65	0.41	0.73 [0.67, 0.79]	-0.08 [-0.12, -0.03]
F-T	158	1517	1061	1309	853	0.64	0.45	0.57 [0.47, 0.67]	-0.11 [-0.20, -0.03]
Random	155	1565	943	1382	760	0.67	0.41	0.80* [0.69, 0.91]	-0.1 [-0.19, -0.03]
T-F	154	1494	842	1468	816	0.65	0.36	0.81* [0.72, 0.91]	-0.02 [-0.08, 0.07]

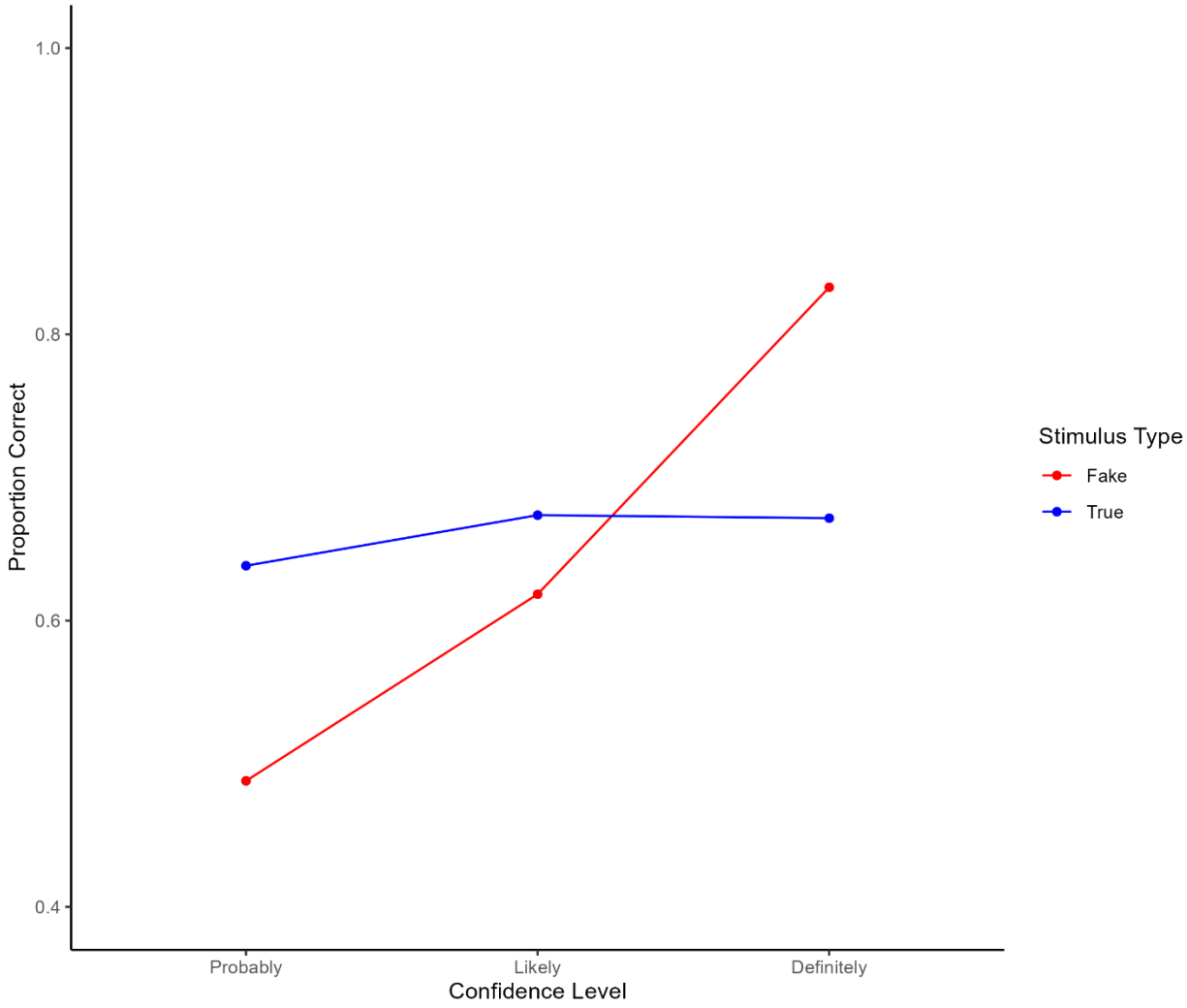
Note. The asterisk indicates that the value was significantly different from F-T condition.

Confidence and Calibration

In addition to evaluating participants' overall accuracy and differences by order, calibration curves were used to assess the relationship between individuals' confidence judgments and proportion correct to determine whether participants who are better calibrated are less susceptible to misinformation. Specifically, I predicted that participants would be better calibrated for true stimuli than fake stimuli (Hypothesis 5). Figure 4 shows the proportion correct for true and fake stimuli across confidence levels (probably, likely, and definitely). As illustrated, calibration for true stimuli appears to be flat relative to fake stimuli. This finding is contrary to my hypothesis. Interestingly, I found that when participants made "definitely" confidence judgements for fake stimuli, they were better calibrated than true stimuli. It is clear from Figure 4 that participants were better calibrated for fake stimuli than true stimuli, especially when rating high confidence.

Figure 4

Calibration Curves for True and Fake Stimulus Type for All Participants



Individual Differences

Mixed effects models were used to control Type I error rates when exploring how individual differences impact discrimination and bias for the news headlines. In the data, participants are nested within experimental conditions of order. But because it is possible that there are other types of order that could impact discrimination and bias, order condition is included as a random effect in the models. Each model explores what individual differences explain participants' confidence judgments, discrimination, and response bias. To make this analysis relevant from a theoretical perspective, I wanted to model which measures were valuable in explaining variability in participants' confidence judgments. Using a random effects ANOVA in mixed effects modeling, it allows us to estimate how much participants vary in their intercepts. This is accomplished by estimating the means based on the clustering variable (i.e.,

participant id number) and is equivalent to an ANOVA because the estimates are based on the means. This is also called a null or unconditional model (Garson, 2013). I also use the interclass correlation coefficient (ICC) to estimate how much of the variability in participant's outcome is explained by the participants themselves. This allows me to compare the reduction in the proportion of unexplained variance from an unconditional model to a model conditional on factors. Although the model predicting confidence judgments includes id as a random factor, for the models predicting discrimination and bias, id cannot be included as a random factor because participants have a single discrimination and bias estimate, which would result in convergence issues. Thus, for the models predicting discrimination and bias, order was included as a random factor. My model predictors (but not order condition and stimulus class) were centered around the grand mean to make the estimates for intercept and slope more interpretable and to reduce multicollinearity (Enders & Tofighi, 2007).

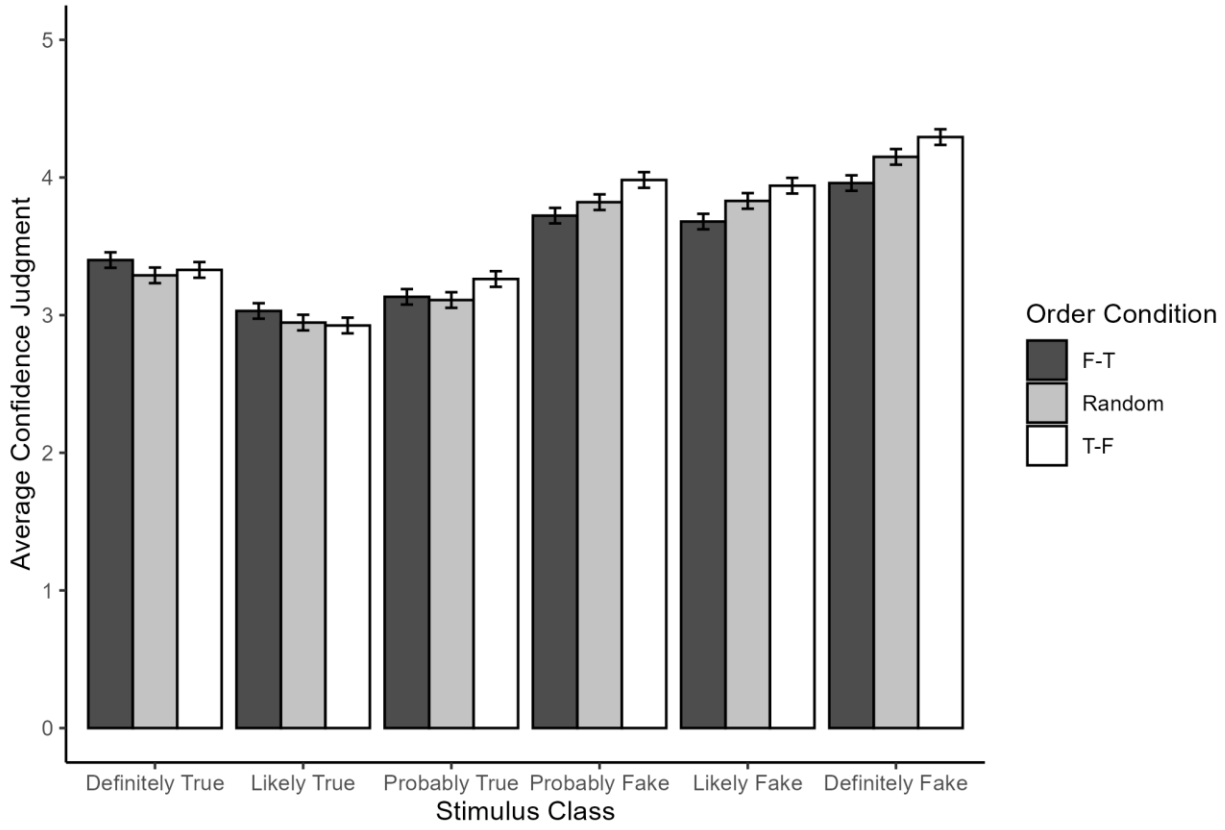
Confidence Judgments

An ICC of 0.2 for the null model estimated that 20% of the residual variance is due to the participants themselves. The random effects ANOVA model estimated the standard deviation for the random effect of id on average confidence judgments to be 0.37. The model found a significant random effect of id, χ^2 (1) = 200.94, $p < 0.05$. This finding indicates that 20% of the residual variability in participant's average confidence judgments is explained by between group differences and 80% is due to individual differences within the groups (id).

Next, I ran the model I am most interested in containing the fixed independent variables (order and stimulus class), individual differences measures, and the random effect of id. This model is equivalent to a random effects ANCOVA. A significant fixed interaction effect of order and stimulus class on confidence judgments was observed, $F(10, 2320) = 4.39$, $p < 0.05$, suggesting that the effect of order on participants' confidence ratings is dependent on the stimulus class, as illustrated in Figure 5. A main effect of stimulus class was found, $F(5, 2320) = 290.47$, $p < 0.05$. However, a main effect of condition was not significant. This finding indicates that when a participant's average confidence judgments is allowed to randomly vary, the effect of condition can change how participants rate their confidence for fact-checked true and fake stimuli that vary in their degree of certainty. This means that participants who are initially exposed to fake stimuli may be induced to mistakenly respond that fake stimuli have more truthfulness, and true stimuli have more falsehood, than other conditions.

Figure 5

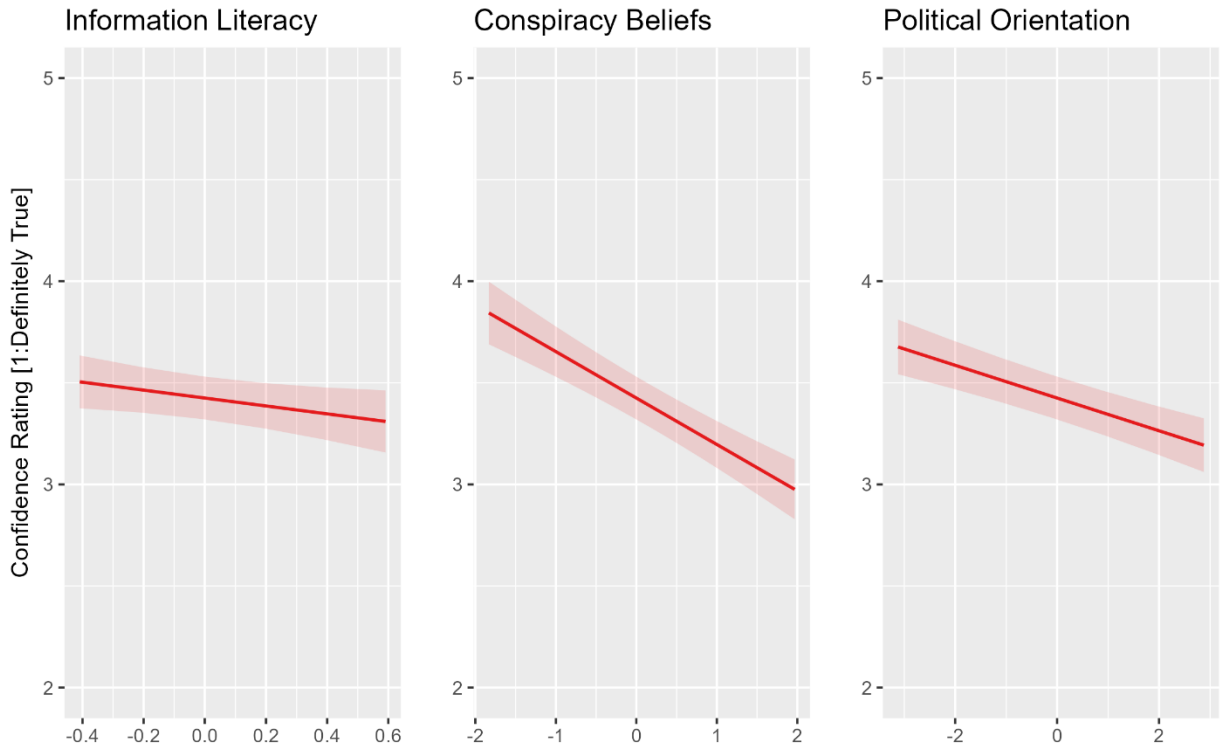
Interaction Effect of Condition and Stimulus Class



Significant fixed effects of information literacy scores, b = - 0.21, se = 0.10, $p < 0.05$, generic conspiracy belief scores, b = -0.23, se = 0.03, $p < 0.05$, and political orientation, b = -0.08, se = 0.01, $p < 0.05$, were found (see Figure 6). These findings suggest that with each unit increase in information literacy scores, generic conspiracy belief scores, and political orientation, participants rated stimuli as 'more true' on average. These results show general support for Hypothesis 6: Confidence judgments differ as a function of individual differences, wherein participants with greater conspiracy beliefs and reporting a conservative political orientation rated stimuli with more truthfulness than participants lower in these measures. Counter to my intuition, I also found the same pattern for participants with greater information literacy. All other measures were not statistically significant. The full ANOVA table for confidence judgments is included in the Appendix.

Figure 6

Average Confidence Rating By ILT, GCBS, and Political Orientation



Note. Confidence rating scale on the y-axis actually ranged from 1:Definitely True to 6: Definitely Fake.

A significant random effect of id was also found, $\chi^2(1) = 318.59$, $p < 0.05$, and estimated that the standard deviation for the random effect of id on average confidence judgments to be 0.34. Including the fixed factors and individual difference measures in the mixed model decreased the proportion of unexplained variance by 38.74%. The proportion of variance explained in the model conditional on both the fixed and random effects was 51.85%. It is clear from this model that average confidence significantly varies across the participants and over half of the variability can be explained when considering order, stimulus class, and individual differences.

Discrimination

The ICC for the null model predicting discrimination found that 4.1% of the proportion of variance in discrimination was due to the order condition. A random effects ANOVA model estimated the standard deviation for the random effect of order on discrimination to be 0.13. The model found a significant random effect of order, $\chi^2(1) = 72.58$, $p < 0.05$. The ICC for discrimination is smaller relative to the confidence model, which simply means there is less dependence in participants' discrimination across order conditions. This is also reflected in the smaller standard deviation. To further clarify, the design effect – the magnitude of adjustment

needed to produce accurate standard error estimates [1+(n – 1)*ICC] – for the discrimination model was 20.3, indicating that I would have needed only 1/20.3 as many participants to achieve this same estimate if a random sample had been collected instead (for a review, see Alimohamadi & Sepandi, 2019; McNeish & Stapleton, 2016). The motivation for a mixed model framework was to better account for sources of variability between and within subjects to achieve less biased estimates. Nonetheless, I found a significant random effect of order, which supports its inclusion.

The random effects ANCOVA model used to predict discrimination included the same variables from the model predicting confidence judgments apart from removing condition and stimulus class because, again, there is only a single estimate of discrimination per participant. Thus, results predicting discrimination and bias do not include interaction terms. The individual difference measures were included as fixed effects. I did not include every possible order direction in my study, so I included it is as a random factor in my discrimination and bias model.

The random effects ANCOVA found significant fixed effects of cognitive reflection scores, b = 0.01, se = 0.05, $p < 0.05$, generic conspiracy belief scores, b = - 0.16, se = 0.02, $p < 0.05$, active open-minded thinking, b = 0.3, se = 0.03, $p < 0.05$, political orientation, b = -0.1, se = 0.01, $p < 0.05$, on discrimination were found (See Figure 7). These results show general support for Hypothesis 2. In particular, I found statistical support for the predictions that discrimination increased as a function of active open-minded thinking and cognitive reflection and decreased as a function of conspiracy beliefs and reporting a more conservative political orientation. This suggests that participants who were more open-minded and reflective in their cognition had better discrimination between true and fake stimuli, whereas those that had higher conspiratorial cognition and reported being conservative had worse discrimination for true and fake stimuli. The full ANOVA table for discrimination is included in the Appendix.

Figure 7

Discrimination Estimates By GCBS, Political Orientation, CRT, and AOT



A significant random effect of condition was observed, $\chi^2(1) = 94.17$, $p < 0.05$. The standard deviation in discrimination between conditions is 0.13. The proportion of variance explained in the model conditional on both the fixed and random effects was 29.17%. Including the fixed factors and individual difference measures in the model decreased the proportion of unexplained variance in participants' discrimination by 26% relative to the null model. These findings indicate that a model allowing the effect of condition to randomly vary, and adjusted for individual differences, could better explain differences in discrimination across conditions than a model that only included a random effect of condition. While a substantial proportion of variability is explained by fixed and random effects, there is still a considerable proportion of variability that remained unexplained by the model.

Bias

The ICC for the null model predicting bias found that 1.47% of the proportion of variance in bias was due to order condition. A design effect of 7.85 was estimated suggesting that the sample variance is over 7 times greater than it would be if I had used a random sample of the same size. A random effects ANOVA model estimated the standard deviation for the random effect of condition on participants response bias to be 0.06. The model found a significant random effect of order, $\chi^2(1) = 22.18$, $p < 0.05$. These results indicate that a mixed model

framework is appropriate for the data and that order has a significant random effect on participant's response bias.

As mentioned above, the model did not include condition or stimulus class as a fixed effect, no interaction term was included, and order condition was included as a random effect. The random effects ANCOVA found significant fixed effects of information literacy, b = - 0.25, se = 0.04, $p < 0.05$, generic conspiracy belief scores, b = - 0.22, se = 0.01, $p < 0.05$, social media usage, b = 0.05, se = 0.02, $p < 0.05$, and political orientation, b = - 0.07, se = 0.01, $p < 0.05$, were found (Figure 8). These results provide general support for Hypothesis 4. I found statistical support for the predictions that response bias decreased as a function of generic conspiracy belief scores and participants who rated their political orientation as more conservative. This suggests that participants who were more conservative, high in conspiratorial cognition, or high in information literacy, relaxed their criteria to rate stimuli as true. Contrary to my prediction, participants with greater social media usage scores were relatively unbiased in their response bias. The full ANOVA table for response bias is included in the Appendix.

Figure 8

Response Bias By ILT, GCBS, Social Media Usage, and Political Orientation



A significant random effect of condition was observed, $\chi^2(1) = 10.44$, $p < 0.05$. The standard deviation in bias between conditions was 0.05. The proportion of variance explained in the model conditional on both the fixed and random effects was 18%. Including the fixed factors and individual difference measures in the model decreased the proportion of unexplained variance in participants' discrimination by 16.76% relative to the null model. Accounting for between group variability of order explained a significant proportion of variance in participants' response bias for true and fake stimuli.

Discussion

My main finding was that, in general, participants were able to discriminate between true and fake news stimuli, but not very well. I found general support for Hypothesis 1, showing that discrimination differs by stimulus class and stimulus type. A curious finding is the decrease when comparing discrimination between headlines labeled as definitely true versus definitely fake. Specifically, when conducting discrimination analyses by different levels of stimulus class, participants had a significant decrease in discriminability for fact-checked stimuli labeled definitely true. This finding suggests that there is something unique about the definitely true stimuli that resulted in participants making more errors and exhibiting a lower overall discrimination ability. I shall have more to say about that in the General Discussion.

I also found evidence that discrimination differs by the order condition, wherein the F-T condition resulted in poorer discrimination relative to other conditions. However, I did not find statistical support for differences in response bias by order, contrary to Hypothesis 3. Participants were less biased in the T-F condition, but it was not statistically different from other order conditions. Instead, the F-T condition resulted in participants having worse discrimination than the T-F and Random condition. One possible explanation for this could be that the F-T order influenced what diagnostic information or features of the headline stimuli participants were willing to consider. It is possible that the F-T condition somehow influenced the quantity or quality of features participants considered and the sequential nature of the study may have interacted with this to result in lower discrimination, especially for the definitely true stimuli. This explanation for order is supported by the eyewitness literature concerned with the design of lineups. When an unfair lineup is used, in which the suspect stands out, it can lead to witnesses focusing on or weighing the wrong features (e.g., generic features like facial hair or race rather than more diagnostic features like nose and eye shape and size), which can impair a witnesses' ability to discriminate between an innocent versus guilty suspect and reduce the accuracy of their judgments compared to using a fair lineup (Colloff et al., 2016). To relate back to my finding, being in the F-T condition may have induced participants to focus more on non-diagnostic features (e.g., the headline source, the image, or social/political factor associated with the statement and/or person) than diagnostic features (e.g., the content of the statement made).

Regarding participants' confidence judgments, I did not find support for Hypothesis 5 that participants would be better calibrated for true than fake stimuli. In fact, participants' accuracy for true stimuli was relatively flat across the confidence levels, and even decreased from 'likely' judgments to 'definitely' judgments. This was an unexpected finding because research has shown that individuals have better accuracy for true headlines than fake headlines

(Pennycook et al., 2020b). These results suggest evidence supporting the alternative: participants were better calibrated for fake stimuli than true stimuli. This finding is consistent with research investigating how accurate are participants' perceptions of social media stimuli with or without a high number of Facebook likes (see Luo et al., 2022). There are several potential psychological explanations for my finding including the possibility that the fake headlines used were just more salient or unusual than true headlines and engaged more of participants' cognitive resources when making judgments. Another possible explanation is that participants have different decision making strategies for verifying true versus fake headlines that may be more effective for fake news accuracy when assigning confidence. The only problem with either of these explanations is it suggests that this should result in changes in response bias, which I did not observe. Because of this, these effects on calibration may better be explained by something unique about the stimuli themselves, especially the definitely true stimuli, a point which will be considered more in the General Discussion.

In my mixed model analyses, confidence judgments did not change by order condition alone; rather, the effect of condition on confidence judgments was dependent on the stimulus class. For true stimuli, participants in the F-T condition generally rated stimuli as more false than other conditions, whereas for the fake stimuli, participants in the F-T condition generally rated stimuli as more truthful. Implications for this finding suggests that subtle changes in order can enhance the effect of the stimuli on confidence judgments. This is an interesting finding that has implications for how individuals make confidence judgments for true and fake headlines, because misinformation could be presented to subjects in a cohesive pattern that 'create' false social narratives as facts and thereby seeks to reinforce individual's beliefs in misinformation. This technique could also be used to operantly condition groups and motivate people to repel alternative explanations and facts.

Differences in individuals' cognition appeared to have an influence on participants' confidence judgments, discrimination, and bias. On average, participants' discrimination ability decreased with greater conspiracy beliefs and participants' rating that they were more conservative. Participants' discrimination ability increased with greater active open-minded thinking and cognitive reflection. This finding is strongly supported by prior research on misinformation and fake news. I also found that participants' response bias was negatively associated with information literacy, conspiracy beliefs, and political orientation. Participants' response bias was positively associated with social media usage. This is an interesting finding because my assumption was that greater social media use would loosen participants' criteria for making a true rating. Participants' confidence judgments decreased with information literacy, conspiracy beliefs, and political orientation. Considering prior research, it may seem intuitive that participants high in conspiracy beliefs, or those that are more conservative, rate their confidence as more true on average. However, the finding suggesting that greater information literacy resulted in participants making more true confidence judgments, is puzzling.

Overall, my results show that participants can discriminate true and fake stimuli, but not very well. Interestingly, participants surprisingly poor ability to discriminate fact-checked stimuli that were labeled as definitely true resulted in lower overall discrimination ability for

participants. In addition, I assumed that participants would be better calibrated in their confidence accuracy when judging true stimuli, but I found that participants were better calibrated for fake stimuli instead. One take-away from Study 1 may be that, for the stimuli used in this study, the truth is harder to identify than falsity. Findings involving what individual differences predict participants' discrimination, bias, and confidence judgments are largely consistent with findings from the literature and suggest that some of these individual differences, like conspiracy beliefs, political orientation, information literacy, cognitive reflection, and social media usage, outperform others.

Study 2

Overview

Study 2 extends study 1 by evaluating whether the relationship of order on discrimination and bias is dependent on whether headlines include a news source. Investigating the influence of news source on aspects of performance is important because research has shown that people rate ideologically congruent headlines as more accurate and are perceived as more credible (Luo et al., 2022; Traberg & van der Linden, 2022). A news source is an additional cue that can provide credibility or reliability to a headline. The general consensus is that the presence of sources in news headlines increases the likelihood of persuasion by adding news credibility, which is intended to protect against misinformation. However, the supporting research is mixed and comes from a relatively small number of studies examining the impact of source credibility on detection accuracy. Moreover, what constitutes a source is not always consistent, with studies using a range of inputs such as a news media company name, a political official, Facebook likes, or an authority agent as sources.

For example, Luo et al (2020) found that headlines with high versus low numbers of Facebook likes (as a proxy for credibility) increased accuracy for true news but decreased accuracy for fake news. Pennycook and Rand (2020) found a negative relationship between analytical reasoning and fake news accuracy, regardless of the news source being present or absent. Research also has found that source credibility influences detection accuracy as a function of analytical reasoning for true (but not fake) news headlines (Pehlivanoglu et al., 2021). Other research has shown there are adverse effects of using sources because they can make participants reject true news or accept fake news based on these other cues. For example, Traberg and van der Linden (2022) found that participants are more accurate at detecting true and fake news from politically congruent sources and perceive them as less slanted and more credible (Batailler et al., 2022). But other research has found no effect of source credibility on accuracy of fake news and suggests that accuracy is due to individual differences (Schaewitz et al., 2020).

The mixed research designs and findings makes it challenging to discern how critical source is to accurately discriminate true and fake news. What is clear is that the inclusion of sources can, in some circumstances, make news headlines appear more credible and accurate, an influence that is likely mitigated by individual differences. From previous research findings, participants with lower analytical reasoning should have lower discrimination with or without a source because they engage is less critical thinking. In addition, participants should be more likely to judge politically congruent headlines as true and politically incongruent headlines as fake.

Participants in study 2 completed the same task with the same stimuli used in study 1 except the fact-checked stimuli included the news source associated with the original post. Participants were randomly assigned to random, true-fake, or fake-true order conditions. Participants then completed the same individual difference measures from study 1. Study 2 also included the same hypotheses and methodology as study 1. While research demonstrates the effect of source is influenced by individual differences, systematically controlling the sources

and making associated predictions was not an aim of this study. Rather, the focus of study 2 was to determine whether I could replicate and extend the findings from study 1 with the inclusion of the source associated with fact-checked stimuli that vary in their degree of truthfulness.

To revisit my hypotheses, I expected discrimination for true and fake stimuli labeled as definitely to be lower than true and fake stimuli labeled as likely or probably. I also expected discrimination to be lower for participants in the fake-true order relative to participants in the true-fake order (Hypothesis 1). I expected response bias to make true or fake judgments to differ by order with participants in the true-fake condition lowering criteria to make a "true" response, participants in the fake-true condition increasing their criteria to make a "true" response to headlines, and participants in the random condition falling somewhere in the middle (Hypothesis 3). For confidence judgments, I evaluated how calibrated participants are in their confidence judgments about true and fake stimuli (Hypothesis 5). Finally, I made predictions about how individual differences in cognition should influence discrimination (Hypothesis 2), response bias (Hypothesis 4), and confidence calibration (Hypothesis 6). In sum, I do not propose different hypotheses in study 2. Rather, I test the same hypotheses but now with the original headline source included to examine how this important factor impacts performance.

Method

Participants

A total of 544 participants completed the study online through the University of Oklahoma's SONA system from 11/21/2023 to 03/20/2024. Data from 44 participants were removed for incomplete data resulting in a total of 480 completed cases for analysis. The majority of participants were between the ages of 18-24 (n = 475, 99%), female (n = 360, 75%), and White/Caucasian (n = 304, 63%). Participants most frequently reported being slightly (n = 169, 35%) or moderately (n = 160, 33%) familiar/knowledgeable about local, national, and world news. Participants most frequently reported their political orientation as unsure or neither conservative nor liberal (n = 114, 24%) or moderately conservative (n = 114, 24%) followed by moderately liberal (n = 81, 17%), slightly conservative (n = 74, 15%), slightly liberal (n = 68, 14%), extremely conservative (n = 17, 3%), and extremely liberal (n = 12, 2.5%). Participants most frequently reported being slightly familiar/knowledgeable (n = 209, 44%) with the study stimuli followed by moderately familiar/knowledgeable (n = 115, 24%), and not at all familiar/knowledgeable (n = 74, 15%).

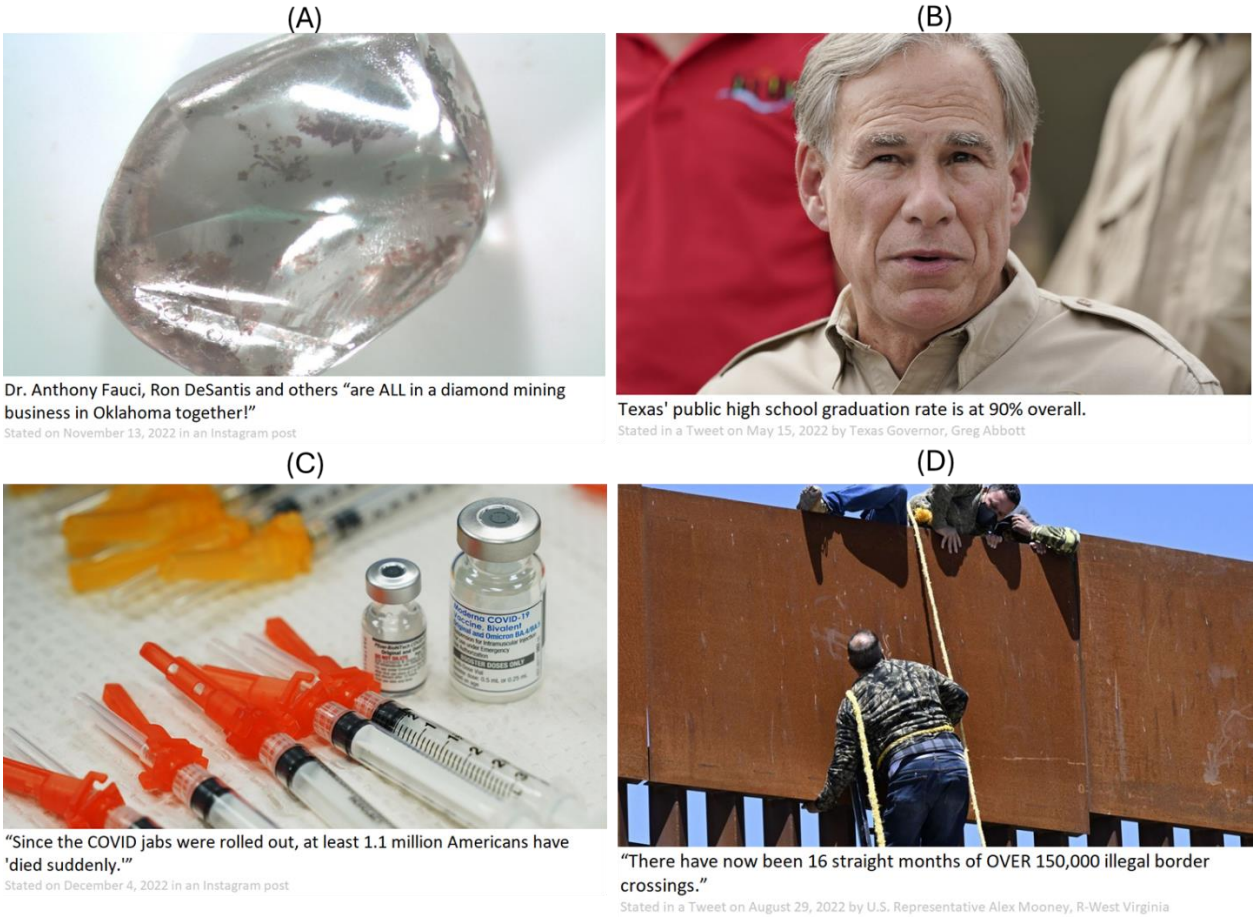
Materials and Stimuli

Fake and True News Stimuli

The same 30 fake and true news stimuli from study 1 were used in study 2. The only difference between study 1 and study 2 stimuli was the inclusion of source in each headline (see Figure 9). The source included in each stimulus was the same source identified by politifact.com. The source included the date and platform the headline was posted and the individual or organization that made the statement. A static image from politifact.com was also included for each stimulus.

Figure 9

Example of Fact-Checked Headline Stimuli with Source Included



Measures

The same measures from study 1 were used in study 2. Participants completed 1) the CRT to measure analytical reasoning, 2) the numeracy skill items, 3) the AOT to measure open minded thinking, 4) the GCB-5 to measure conspiratorial cognition, and 5) the ILT to measure information literacy. A complete list of the measures' items is included in the Appendix.

Design

The study 2 design was the same as study 1. The study design was a 6 stimulus type (Definitely Fake, Likely Fake, Probably Fake, Probably True, Likely True, and Definitely True) x 3 stimulus order (True-Fake, Fake-True, or Random). Stimulus type was a within-subjects factor and stimulus order was a between-subjects factor. Participants were randomly assigned to an order condition.

Procedure

The procedure in study 2 was the same as study 1. Participants completed an informed consent form followed by the MIST as practice trials. During the experimental trial, participants were randomly assigned to order condition (True-Fake, Fake-True, or Random) and instructed to rate how likely they would be to share the headline on social media on a 1-6 Likert scale (Not at all likely-Extremely likely) and rate how confident they were each headlines was fake or true on a 1-6 Likert scale (Definitely Fake-Definitely True). After completing the experimental trials, participants rated how familiar they were with any of headlines and their familiarity/knowledge for various news-related domains. Next, participants completed the CRT and numeracy skills items in random order followed by the GCB-5, ILT, and AOT in a random order. Finally, participants rated their social media use across various platforms and completed demographic items. Finally, participants were presented with a debriefing page that informed them they completed the study and would receive course credit.

Results

Outlier Detection

Data were screened for outliers and adjusted in the same way as study 1:

1. Participants that had only used one side of the scale (responded true or fake for all stimuli) were removed. This resulted in the removal of data from 2 participants.
2. Boxplots of discrimination ability (*d'*; described below) identified 9 outliers that were removed.

The removal of these 11 outliers resulted in a total of 469 cases for analysis. In alignment with study 1, participants with negative *d'* estimates were retained.

Descriptives

Factors

Grouped by order, an ANOVA test found no significant group mean differences in participants' confidence judgments. An ANOVA found significant differences between the proportion of correct responding across conditions, $F(2, 466) = 4.32$, $p < 0.05$, $\eta^2 = 0.02$. Post-hoc comparisons using Tukey's HSD test found that the proportion of correct responding was significantly lower in the F-T ($M = 0.61$, $SD = 0.49$) condition relative to both the T-F ($M = 0.64$, $SD = 0.48$) and Random ($M = 0.64$, $SD = 0.48$) condition.

Grouped by stimulus type, a paired sample *t*-test found that participants had significantly different mean confidence judgments, $t(468) = 25.38$, $p < 0.05$, $d = 1.17$, with confidence for fake ($M = 3.89$, $SD = 1.36$) stimuli being rated statistically higher, on average, compared to confidence for true ($M = 3.14$, $SD = 1.20$) stimuli. A paired sample *t*-test found that participants had significantly different rates of correct responding, $t(468) = 5.86$, $p < 0.05$, $d = 0.27$, with the proportion of correct responding being higher for true ($M = 0.67$, $SD = 0.47$) stimuli compared to fake ($M = 0.59$, $SD = 0.49$) stimuli on average.

Grouped by stimulus class, an ANOVA found significant group mean differences in participants' confidence judgments, $F(5, 2808) = 178.4$, $p < 0.05$, $\eta^2 = 0.24$. Post hoc comparisons using Tukey's HSD test found significant differences between all comparisons except when comparing probably fake ($M = 3.78$, $SD = 1.33$) and likely fake ($M = 3.83$, $SD = 1.37$) stimuli. An ANOVA test found significant differences between participants' rate of correct responding, $F(5, 2808) = 44.9$, $p < 0.05$, $\eta^2 = 0.07$. Post hoc comparisons using Tukey's HSD test found significant differences for all comparisons except probably fake and definitely true, likely fake and definitely true, definitely fake and probably true, and likely fake and probably fake (see Table 6).

Table 6

Sample Size (n), Average Confidence (SD), p(Correct) (SD), and Average Sharing (SD) By Condition, Stimulus Type, and Stimulus Class

Condition	*n*	Average Confidence (SD)	p(Correct) (SD)	Average Sharing[4] (SD)
F-T	155	3.54 (1.35)	0.61 (0.49)	1.69 (1.22)
Random	155	3.47 (1.29)	0.64* (0.48)	1.74 (1.18)
T-F	159	3.54 (1.37)	0.64* (0.48)	1.71 (1.20)

Note. The asterisk indicates that the value was significantly different from the F-T condition.

Stimulus Type	*n*	Average Confidence (SD)	p(Correct) (SD)	Average Sharing[5] (SD)
True	469	3.14 (1.20)	0.67 (0.47)	1.82 (1.28)
Fake	469	3.89* (1.36)	0.59* (0.49)	1.60* (1.11)

Note. The asterisk indicates that the value was significantly different from True stimuli.

	Stimulus Class	*n*	Average Confidence (SD)	p(Correct) (SD)	Average Sharing[6] (SD)
1	Definitely True	469	3.36 (1.26)	0.59 (0.49)	1.82 (1.28)
2	Likely True	469	2.93* (1.07)	0.75* (0.43)	1.82 (1.28)
3	Probably True	469	3.14* (1.21)	0.67* (0.47)	1.83 (1.28)
4	Probably Fake	469	3.83* (1.37)	0.56 (0.5)	1.60* (1.12)
5	Likely Fake	469	3.78* (1.33)	0.56 (0.5)	1.61* (1.11)
6	Definitely Fake	469	4.07* (1.38)	0.64* (0.48)	1.57* (1.09)

Note. The asterisk indicates that the value was significantly different from Definitely True stimuli.

[4] No differences were found in participants' sharing judgments by order.

[5] A paired sample *t*-test found significant different mean sharing judgments, $t(468) = 9.96$, $p < 0.05$, $d = 0.46$, with likelihood to share being higher for true ($M = 1.82$, $SD = 1.28$) stimuli compared to fake ($M = 1.60$, $SD = 1.11$) stimuli.

[6] An ANOVA found significant group mean differences in participants' likelihood to share across stimuli class, $F(5, 2808) = 8.6$, $p < 0.05$, $\eta 2 = 0.02$. Post hoc comparisons using Tukey's HSD test found significant differences between definitely true and definitely fake, likely true and definitely fake, probably true and definitely fake, likely fake and definitely true, probably fake and definitely true, likely true and likely fake, probably true and likely fake, probably fake and likely true, and probably true and probably fake.

Individual Differences

Descriptive statistics for the individual difference measures are shown in Table 7. Accuracy on the cognitive reflection test, numeracy skills, and information literacy test were generally at or below 50/50 chance. Participants had similar scores of GCB-5 and AOT to study 1. I also found similar scores of social media usage. On average, participants rated being unsure or neither liberal nor conservative political orientations (n = 114), followed by slightly, moderately, or extremely conservative (n = 205) and slightly, moderately, or extremely liberal (n = 161). Most participants were slightly to moderately familiar with and knowledgeable about local, national, and world news.

Table 7

Sample Size (n), Mean, Standard Deviation (SD), and Median by Individual Difference Measure

Measure	*n*	Mean	SD	Median
Cognitive Reflection Test	469	0.44	0.26	0.43
Numeracy Skills	469	0.55	0.25	0.5
Information Literacy Test	469	0.40	0.23	0.4
Generic Conspiracist Beliefs	469	2.89	0.73	3
Active Open-Minded Thinking	469	3.72	0.42	3.7
Media Use Composite	469	2.28	0.40	2.23
Political Orientation	469	4.18	1.56	4
News Knowledge	469	2.86	0.99	3

Discrimination and Response Bias

As done in study 1, signal detection measures (discrimination and response bias) are reported at the item and subject level. Response rates were also adjusted for perfect (N-1) or zero discrimination (1/N).

Item level

The *d'* adjacent and *d'* cumulative were estimated to examine how well participants could discriminate between the classes of stimuli (see Table 8). Interestingly, as in study 1, a negative *d'* adjacent was observed for definitely true stimuli relative to likely true stimuli. This suggests that participants perceived the definitely true stimuli as more fake than they actually were relative to their perceptions of likely true stimuli. This resulted in a similar lower overall cumulative discrimination ability between definitely fake to definitely true stimuli (d' = 0.60) as found in study 1. Again, I found that, despite the trouble with the definitely true stimuli, participants could reasonably discriminate between the true and fake news headlines. The probability of responding true (i.e., p(True)) illustrates the same large decrease in the proportion of making a true response for definitely true stimuli as found in study 1 (see Figure 10).

The summarized variables described thus far tell a very similar story to study 1. That is, participants can discriminate between true and fake, but not very well. The lower overall

discrimination was due to participant's perceptions of the definitely true stimuli. Next, I will evaluate the same hypotheses for participants in study 2. Specifically, I test whether discrimination for fact-checked headlines that are labeled as definitely true and definitely fake is larger than those that are labeled as likely or probably true and fake (Hypothesis 1), and whether response bias differs by order (Hypothesis 3) in the Discrimination and Response Bias sections. In addition, I evaluate how calibrated study participants are in their confidence judgments about true and fake stimuli (Hypothesis 5) in the Confidence and Calibration section. Finally, I explore what factors may explain individuals' discrimination (Hypothesis 2), response bias (Hypothesis 4), and confidence judgments (Hypothesis 6) in the Individual Differences section.

Table 8

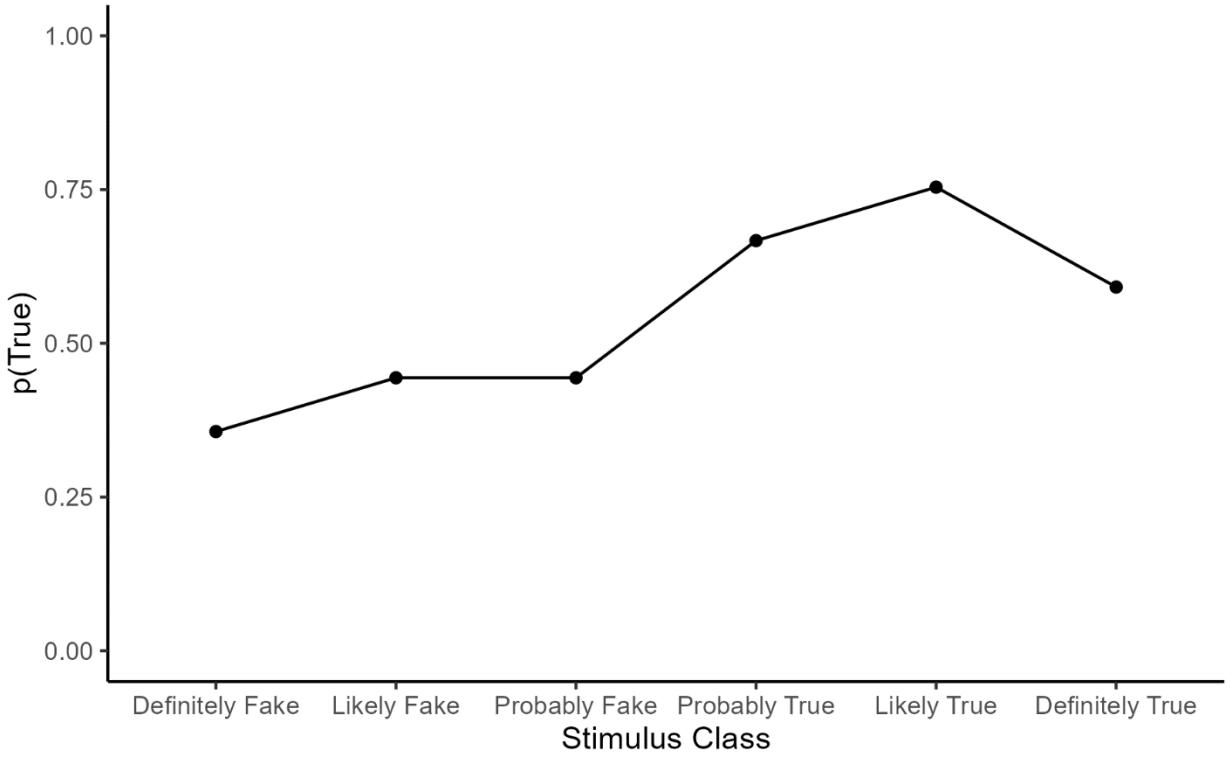
Probability of Responding Fake and True, d' Adjacent, and d' Cumulative by Stimulus Class

	Stimulus Class	p(Fake)	p(True)	d' Adjacent	d' Cumulative
1	Definitely Fake	0.643	0.357	NA	NA
2	Likely Fake	0.556	0.444	0.227	0.227
3	Probably Fake	0.556	0.444	0.000	0.227
4	Probably True	0.333	0.667	0.573	0.800
5	Likely True	0.246	0.754	0.255	1.055
6	Definitely True	0.409	0.591	-0.456	0.599

Note. $n = 480$.

Figure 10

Probability of Responding True Across Stimulus Class



I assessed how discrimination changed across congruent stimuli pairs and whether participants were better at discriminating headlines that were fact-checked as definitely true and fake versus headlines that were less certain in their veracity (Hypothesis 1). Results found that discrimination for likely fake and true stimuli achieved the highest discrimination (d' = 0.83) and was statistically larger than definitely fake and true stimuli (d' = 0.60), $t(468) = 14.63$, $p < 0.05$, and probably fake and true stimuli (d' = 0.57), $t(468) = 13.14$, $p < 0.05$. However, definitely and probably fake and true stimuli were not significantly different from each other (see Table 9). I found minor support again for my first hypothesis that discrimination should increase with certainty in fact-checked headlines. I replicated the unexpected finding that discrimination was worse for definitely true stimuli, which resulted in poorer discrimination when comparing the stimulus class pairs. In other words, participants perceived headlines fact-checked as definitely true as fake more often than they should have.

Table 9

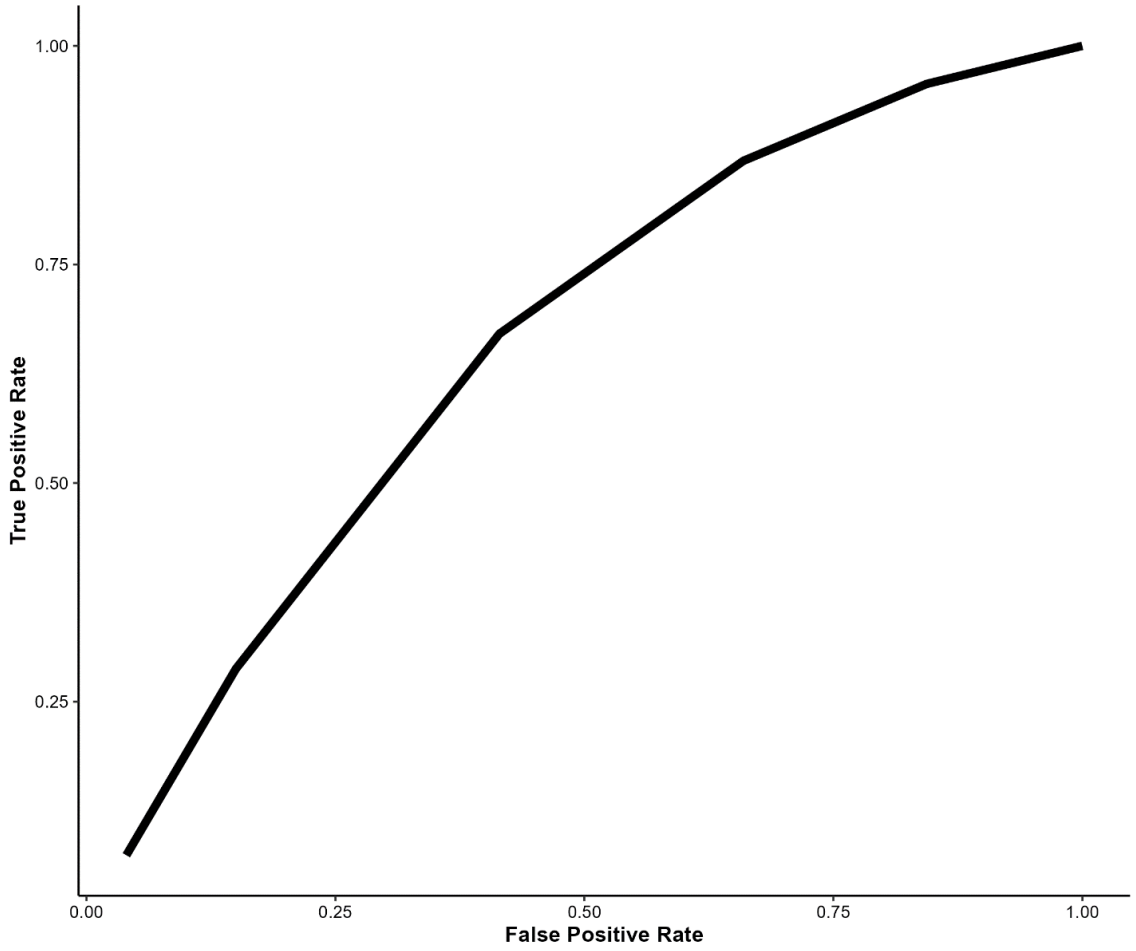
Discrimination (d') Estimates by Stimulus Class Pairs

Stimulus Class	*d'*
Definitely Fake vs Definitely True	0.599
Likely Fake vs Likely True	0.828
Probably Fake vs Probably True	0.573

ROC curves show more similarities in participants' discrimination for the headlines despite the headline source being included in study 2 with an estimated AUC of 0.63 (see Figure 11). The AUC was significantly above chance, $t(468) = 5.59$, $p < 0.05$, but was not significantly different from the AUC in study 1. These findings indicate that even when source was included with the fact-checked headline stimuli, participants were relatively poor classifiers.

Figure 11

ROC Curve for All Participants



Subject level

The response types, response rates, and signal detection measures (*d'* and bias) for the overall sample (n = 469), separated by order conditions, was estimated (see Table 10). Overall, participants could discriminate between true and fake news, but not much better relative to study 1 (*d'* = 0.66) and were slightly biased to respond fake to stimuli (c = -0.11).

An ANOVA test found significant group mean differences in discrimination across condition, $F(2, 466) = 3.08$, $p < 0.05$, $\eta^2 = 0.01$. However, post hoc tests using Tukey's HSD test could not draw confident conclusions regarding which pairs are different and could not meet significance. This is likely due to the marginally significant finding for the ANOVA test (p = 0.047). Order may influence participants' discrimination (Hypothesis 1), but the effect was not large. An ANOVA was performed to address whether response bias differed by order (Hypothesis 3) but found no significant differences.

Table 10

Counts of Participants Response Types, Response Rates, Sensitivity (d'), and Response Bias (*c*)

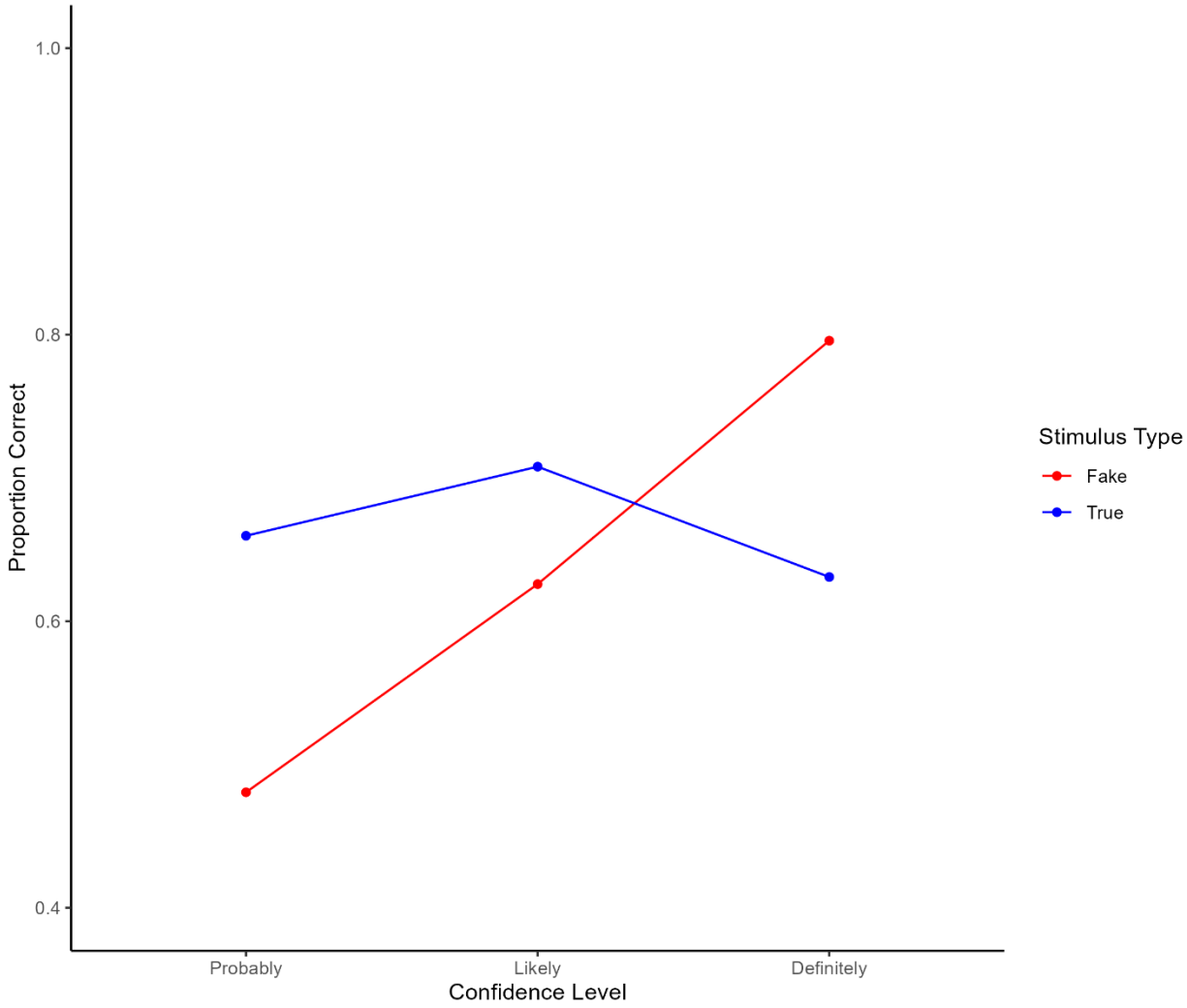
	n	Hits	FAs	CRs	Misses	HR	FAR	*d'* [95% CI]	*c* [95% CI]
Overall	469	4719	2918	4117	2316	0.67	0.41	0.76 [0.70, 0.82]	-0.11 [-0.16, -0.07]
F-T	155	1497	988	1327	828	0.64	0.43	0.65 [0.55, 0.75]	-0.10 [-0.18, -0.01]
Random	155	1626	980	1345	699	0.70	0.42	0.81 [0.71, 0.91]	-0.16 [-0.25, -0.1]
T-F	159	1596	940	1445	789	0.67	0.39	0.81 [0.7, 0.92]	-0.08 [-0.15, -0.00]

Confidence and Calibration

Calibration curves were estimated to examine whether participants were better calibrated for true or fake stimuli (Hypothesis 5). Figure 12 shows the proportion correct for true and fake stimuli across confidence levels (probably, likely, and definitely). Participants appeared to be relatively calibrated for fake stimuli, but there is a different story for true stimuli. True stimuli appeared to have a noticeable increase from probably true to likely true, but participants had a large decrease in accuracy when rating stimuli as definitely true. This finding is consistent with what I found in study 1 and is even more pronounced in study 2.

Figure 12

Calibration Curves for True and Fake Stimulus Type for All Participants



Individual Differences

Mixed models were used to explore the influence of individual differences on aspects of participants' performance. As was done in study 1, when modeling participants' confidence judgment, id was included as a random factor, and condition was included as a random factor when modeling participants' discrimination and response bias. Model predictors other than order and stimulus class factors were centered around the grand mean.

Confidence Judgments

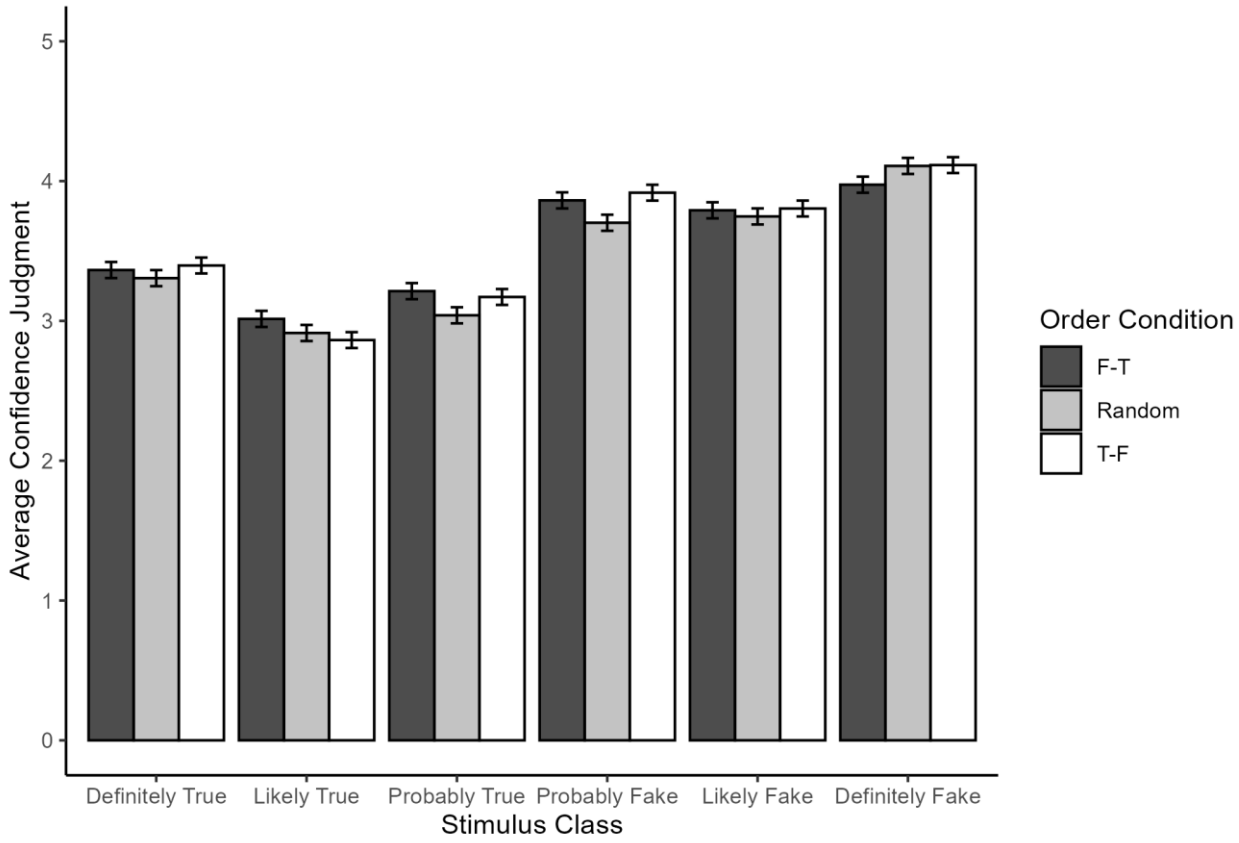
An ICC of 0.24 was estimated for the null model suggesting that 24% of the residual variance in participants' confidence judgments was due to the participants themselves. The random effects ANOVA model estimated the average confidence judgement deviated 0.40 across

participants. The model found a significant random effect of id, $\chi^2(1) = 273.96$, $p < 0.05$. The null model suggests that a sizeable portion of the residual variability in participants' average confidence judgments is explained by between group differences.

A random effects ANCOVA model found a significant fixed interaction effect of order and stimulus class on confidence judgments, $F(10, 2330) = 2.52$, $p < 0.05$ (see Figure 13). This finding suggests that how participants rate their confidence across stimulus class depends on condition. I found a main effect of stimulus class, $F(5, 2330) = 288.96$, $p < 0.05$. However, I did not find a main effect of condition. This finding replicated what I found in study 1: When participants' average confidence judgments were allowed to randomly vary, the effect of stimulus class is compounded by condition. This means that participants' average confidence judgments change with condition when considering varying degrees of certainty in fact-checked true and fake headlines. For instance, participants in the fake-true condition rated definitely fake stimuli as more truthful and rated likely and probably true stimuli as more false relative to other conditions. Conversely, differences in average confidence judgment for definitely true stimuli do not appear to be large.

Figure 13

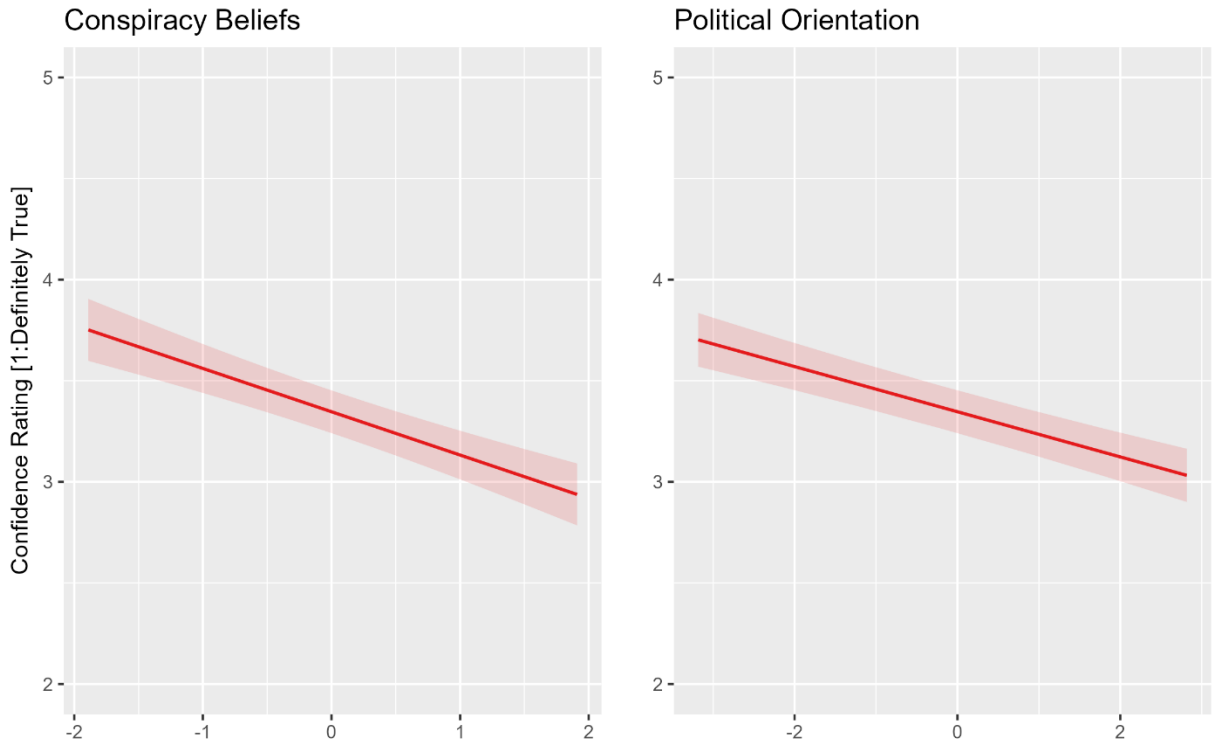
Interaction Effect of Condition and Stimulus Class



I also found significant fixed effects of generic conspiracy belief scores, b = - 0.21, se = 0.03, $p < 0.05$, and political orientation, b = - 0.11, se = 0.03, $p < 0.05$ (see Figure 14). These findings suggest that as belief in conspiracies increased, and participants reported being more conservative, participants rated stimuli as more truthful on average. This result provides general support for Hypothesis 6 that confidence judgments differ as a function of individual differences. For instance, participants with greater belief in conspiracies and who reported being more conservative rated stimuli as more truthful on average than participants with lesser belief in conspiracies and who reported being more liberal. The full ANOVA table for confidence judgments is included in the Appendix.

Figure 14

Average Confidence Rating By GCBS, and Political Orientation



Note. Confidence rating scale on the y-axis ranged from 1:Definitely True to 6: Definitely Fake.

A significant random effect of id was also found, $\chi^2(1) = 370.89$, $p < 0.05$, and estimated that the standard deviation for the random effect of id on average confidence judgments to be 0.36. Including the fixed factors and individual difference measures in the mixed model decreased the proportion of unexplained variance by 38.33%. The proportion of variance explained in the model conditional on both the fixed and random effects was 53.55%. This model demonstrates confidence judgments are explained by interactions between order condition and stimulus class. Moreover, significant effects of individual differences including belief in

conspiracies and political orientation showed a negative relationship with average confidence judgments. Over half of the model variability could be explained when considering order, stimulus class, and individual differences.

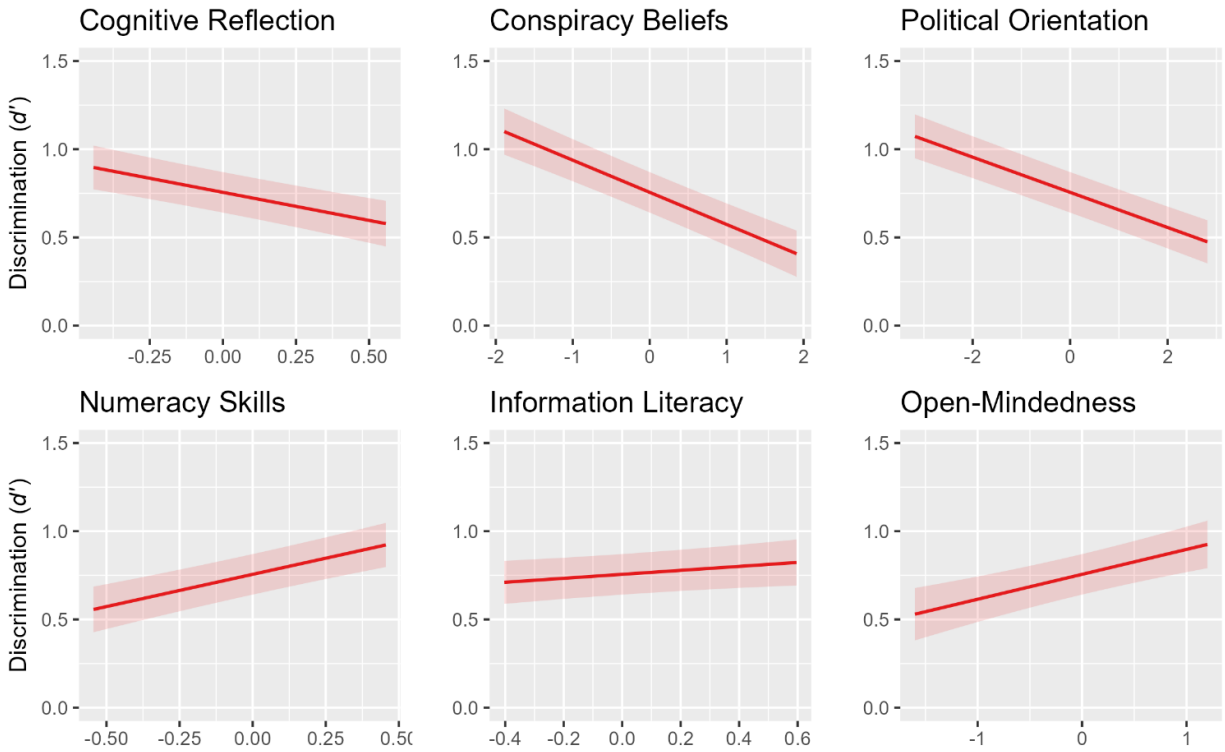
Discrimination

The ICC for the discrimination null model found that only 1.84% of the proportion of variance in discrimination was due to order condition. A random effects ANOVA model estimated the standard deviation for the random effect of order on discrimination to be 0.09. The null model found a significant random effect of order, $\chi^2(1) = 29.05$, $p < 0.05$. This ICC is smaller than anticipated. However, I also found a design effect of 9.61, which supports the use of order being used as a random effect in the model.

The random effects ANCOVA found several significant fixed effects on discrimination. Participants with higher cognitive reflection, b = - 0.32, se = 0.05, $p < 0.05$, greater belief in conspiracies, b = - 0.18, se = 0.02, $p < 0.05$, and who reported being more conservative, b = - 0.1, se = 0.03, $p < 0.05$, had significantly lower discrimination that those with lower scores or reporting being more liberal. Participants with higher numeracy skills, b = 0.37, se = 0.06, $p < 0.05$, information literacy scores, b = 0.11, se = 0.05, $p < 0.05$, and open-mindedness, b = 0.14, se = 0.03, $p < 0.05$, had significantly higher discrimination than those with lower scores (see Figure 15). These results show support for Hypothesis 2 that individual differences influence discrimination. This finding suggests a significant negative relationship between cognitive reflection, conspiracy beliefs, and political orientation on discrimination. Further, I also found a significant positive relationship between numeracy skills, information literacy, and actively open-minded thinking on discrimination. These findings are consistent with what I found in study 1, except for cognitive reflection. In study 1, I found a significant positive effect of cognitive reflection on discrimination, but study 2 showed a significant negative relationship on discrimination. The full ANOVA table for discrimination is included in the Appendix.

Figure 15

Discrimination Estimates By CRT, GCBS, Political Orientation, Numeracy Skills, ILT, and AOT



A significant random effect of condition was observed, $\chi^2(1) = 44.42$, $p < 0.05$. The standard deviation in discrimination between conditions is 0.1. The proportion of variance explained in the model conditional on both the fixed and random effects was 19.82%. Including the fixed factors and individual difference measures in the model decreased the proportion of unexplained variance in participants' discrimination by 17.72% relative to the null model. Findings suggest that a model allowing the effect of condition to randomly vary, adjusted for individual differences, could better explain differences in discrimination, than a model that only included a random effect of condition. This is interesting because it suggests that the inclusion of individual differences in my model contributes to explaining how participants differ in their discrimination.

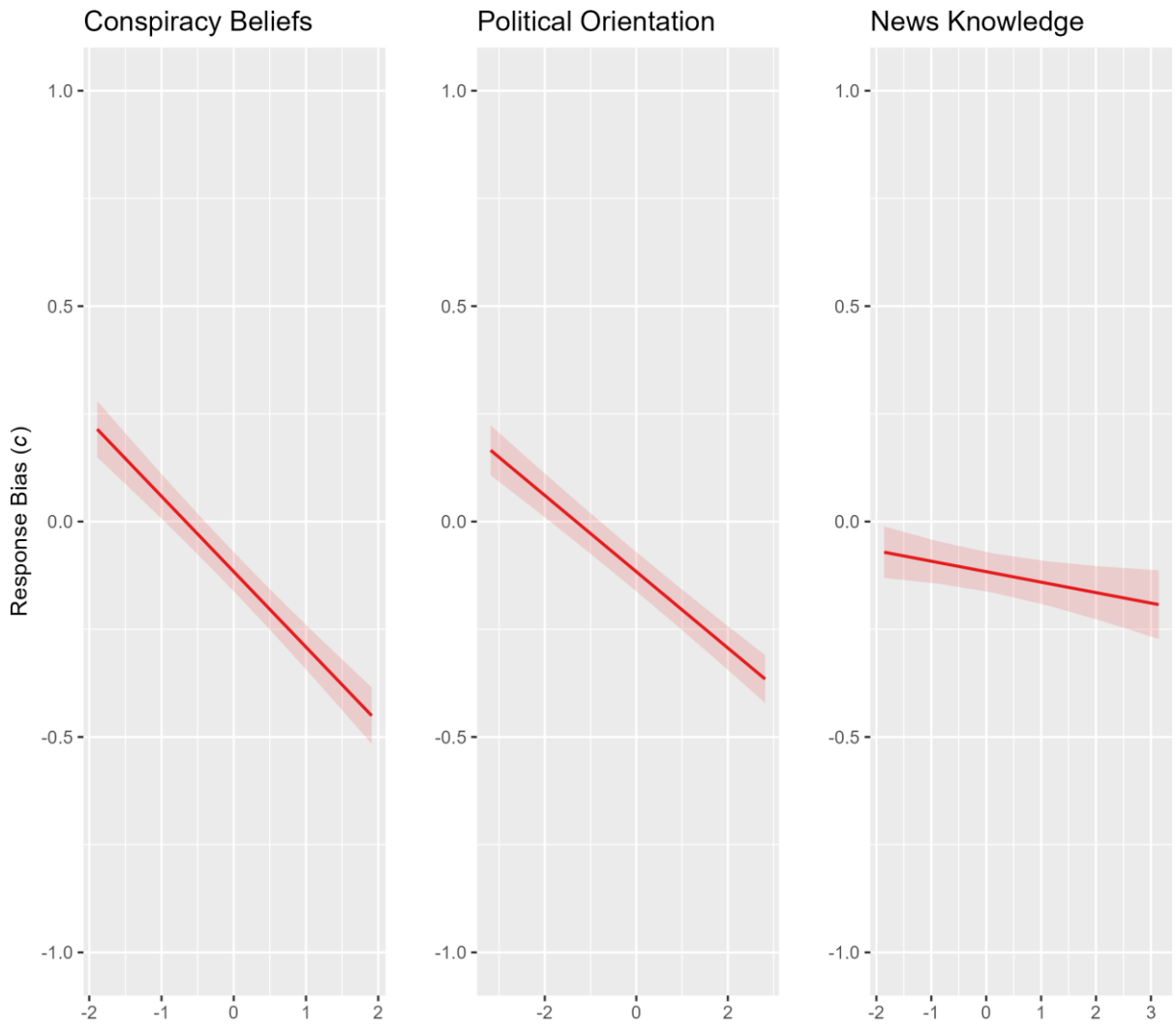
Bias

The ICC for the null model predicting bias found that 0.92% of the proportion of variance in bias was due to order condition. A design effect of 5.31 was estimated suggesting that the sample variance is over 5 times greater than it would be if I had used a random sample of the same size. The random effects ANOVA model estimated the standard deviation for the random effect of condition on response bias to be 0.05. The null model also found a significant random effect of order, $\chi^2(1) = 12.88$, $p < 0.05$. Findings from this model suggest that a mixed model framework is appropriate for the data and that order has a significant random effect on response bias.

The random effects ANCOVA model found significant fixed effects of generic conspiracy belief scores, b = - 0.18, se = 0.01, $p < 0.05$, political orientation, b = - 0.09, se = 0.01, $p < 0.05$, and news knowledge, b = - 0.02, se = 0.01, $p < 0.05$ (see Figure 16). These findings provide general support for Hypothesis 4 that bias should differ as a function of the individual difference measures. Further, I found statistical support that response bias would decrease as a function of conspiracy beliefs and reporting a conservative ideology. I also found support that participants who reported having more news knowledge had a tendency to respond fake to stimuli versus those with less news knowledge. The full ANOVA table for response bias is included in the Appendix.

Figure 16

Response Bias By GCBS, Political Orientation, and News Knowledge



A significant random effect of condition was observed in my full model, $\chi^2(1) = 9.1$, $p < 0.05$. The standard deviation in bias between order conditions was 0.04. The proportion variance explained in the model condition on both the fixed and random effects was 19%. Including the fixed factors and individual difference measures in the model decreased the proportion of unexplained variance in participants' discrimination by 18.3% relative to the null model. Accounting for additional sources of variability by allowing order to randomly vary explained a significant portion of variance in participants' tendencies to respond true or fake to stimuli.

Discussion

The purpose of study 2 was to replicate and extend the findings from study 1 by including the source of the headline stimuli. The main takeaway is that I found more similarities between study 2 and study 1 than differences. I found support for Hypothesis 1 that discrimination differs when comparing stimulus class pairs that matched in fact-checked labels (e.g., definitely, likely, and probably), with discrimination for definitely true and fake stimuli being much lower than expected. I also found the same decrease in discrimination when comparing definitely true versus definitely fake stimuli as in study 1.

I did not find support for Hypothesis 3 predicting response bias differed by order condition. I found that participants in the T-F condition were least biased in their responses but were not significantly different from other conditions. An ANOVA test found evidence that discrimination differed by order. Discrimination was lowest for the F-T condition and discrimination was relatively the same for both the Random and T-F condition. However, post-hoc tests could not pinpoint significant differences in discrimination across order conditions.

I did not find support for Hypothesis 5 that participants would be better calibrated for true than fake stimuli. As in study 1, accuracy for true stimuli had a dramatic decrease in discriminating fact-checked stimuli labeled as definitely true but not definitely fake. This is puzzling because participants should be most accurate when making highly confident decisions, as was the case for definitely fake stimuli in study 1 and study 2. In addition, as mentioned in the study 1 Discussion, research suggests individuals have better accuracy for true versus fake headlines (Pennycook et al., 2020b). Instead, I found that participants were better calibrated for fake stimuli and accuracy increased across their confidence levels. This finding could be partially explained by the poor discrimination found for definitely true stimuli. However, my findings are also supported by prior research findings that participants more accurately identify fake news than true news in a social media context (Luo et al., 2022).

I found an interaction effect of order and stimulus class on participants' confidence judgments showing that for many stimulus classes, participants in the F-T condition rated fake stimuli as more true (as for definitely fake) or rated true stimuli as more fake (as for likely and probably true), on average, relative to other order conditions. It appears that this interaction is driven by the effect of stimulus class rather than order because I only found a significant main effect of stimulus class in my mixed model. This finding was similar to study 1 but did not match the exact pattern of order effects on confidence judgments as in study 2. Nonetheless, this finding suggests that order does have an impact when headline stimuli that are graded in their veracity are considered. This finding suggests that order and stimulus class are factors to consider when

investigating how individuals make judgments about news headlines. Furthermore, I found that the influence of stimulus class and order together is present, whether a source is included or not.

Individual differences also had influences on aspects of participants' performance. On average, discrimination for true versus fake stimuli significantly decreased with cognitive reflection and conspiracy beliefs, but significantly increased with numeracy skills, information literacy, and active open mindedness. Study 2 found that CRT had a significant negative effect on discrimination, which is interesting because this is inconsistent with my study 1 findings. A cognitive analytical account of misinformation identification assumes a positive relationship between CRT and discrimination. Further, a motivational account assumes that an individual higher in CRT may have more resources to strategically process information in a manner that result in inferences that are congruent with their prior beliefs (see Batailler et al., 2022). From this it appears that cognitive reflection, in addition to other individual differences including belief in conspiracies, political orientation, numeracy skills, information literacy, and open mindedness, have an impact on participants' discrimination. While my study 1 results find support for the cognitive account, my study 2 results find support for the motivational account for discriminating. However, the main takeaway from my mixed model analyses is that participants' belief in conspiracies, and those reporting being more conservative, consistently impacted all aspects of performance across both studies.

It is also interesting that I did not find large changes in overall estimates when source was included. Previous research has shown that including a source can increase credibility which should impact accuracy. Other research has found that manipulating credibility may increase accuracy (Luo et al., 2022; Traberg et al., 2024), but at the expense of making different types of errors, such as errors of commission (false alarm) or errors of omission (misses). For instance, errors of commission may be more likely to occur when confronted with headlines that are politically congruent or in favor of a belief while errors of omission may occur when confronted with headlines are politically incongruent or in opposition of a prior belief. These same errors could also be influenced by factors such as who made the statement. One explanation for not observing large differences with headline source included could be due to having participants make sharing and confidence judgments together. This methodology was chosen to influence accuracy, but it may have had a stronger impact on aspects of performance than any potential beneficial effects of source inclusion. It is also possible that the fact-checked stimuli used in this study were challenging or ambiguous to an extent that the source information could not overcome.

An alternative explanation could be due to the context of these types of decisions. Social media headlines are often attention grabbing to attract clicks and may be, as a consequence, perceived as less factual or credible in general. In the present study, confidence judgments were related to participants' own perceptions of the stimuli and participants may have had little motivation to be accurate to a degree of certainty. Moreover, participants did not appear to change their decision making much when a source was included. It is also possible that headline source may have not been attended to at all by some participants. Nonetheless, information source is relevant to decision making and participants may weigh source information more

deeply when making decisions with real-world consequences like voting or affirming their beliefs in front of people they care about.

General Discussion

The focus of this research was to investigate how participants handle fact-checked true and fake headlines. I evaluated how stimulus order (true-fake, fake-true, or random), stimulus class (definitely, likely, and probably true and fake headlines), and individual differences influence aspects of performance. I proposed hypotheses related to how these factors would impact participants' discrimination, response bias, and confidence judgments. The main takeaway from my research is manifold: 1) participants could discriminate true and fake headlines, but not with high accuracy, with surprisingly poor performance found for fact-checked headlines labeled as definitely true, 2) participants were better calibrated in the accuracy of their confidence judgments for fake headlines relative to true headlines, and 3) belief in conspiracies and political orientation, in particular, predicted all aspects of performance (confidence judgments, discrimination, and response bias). These results were consistent across two studies, whether the headline source was not (study 1) or was included (study 2).

I found several interesting findings that deserve more attention. To begin, in both studies I observed a large decrease in the proportion of participants making a true response when the stimuli were fact-checked as being definitely true. There are several possible reasons for this that future research should explore.

One possible explanation could be that when presented with the definitely true stimuli, participants may experience cognitive dissonance with their prior beliefs or attitudes and may be, as a result, less likely to judge the headlines labeled as definitely true in this study as true. This might be because the stimuli may have confused participants, for whatever reason, and they did not perceive them as corresponding to their fact-checked labels. Similarly, participants may simply be more unwilling to make true judgments with the highest confidence than they are to make fake judgments with the highest confidence. This could be because they are operating under the assumption that there is more misinformation online than correct information.

Another explanation for this interesting finding could be due to something about the selected stimuli. The true and fake headlines were normed for each stimulus class using volunteers' judgments and corroborated by ChatGPT. However, participants in both studies showed a large decrease for the definitely true stimuli suggesting that my norming sample may not have been representative of the participant population sampled in my studies. Alternatively, there could be something unique about the definitely true stimuli that resulted in participants making lower confidence judgments and having worse overall discrimination. Definitely true stimuli may have involved features that may have been incongruent with the participant's beliefs, resulting in them rejecting a statement as true because of a political figure that made the statement or when confronted with facts about an emotionally charged topic that conflict with their anchored beliefs.

I also found that participants were better calibrated for fake stimuli than true stimuli. This is also apparent when comparing discrimination for definitely true and fake headlines relative to

the likely and probably stimulus class pairs (see Figure 4 and Figure 12). One explanation for this interesting and surprising result could be due to the methodology I used. Presenting both sharing and accuracy judgments together was chosen to impact discrimination, but it may have had a stronger influence on discrimination for definitely true stimuli than other stimulus classes. This logic comes from previous research showing that participants were worse at discriminating truth from falsehood when evaluating sharing first followed by accuracy judgments than assessing accuracy alone (Epstein et al., 2023). I used a similar methodology in my study to have a stronger impact on discrimination by having participants rate sharing and accuracy simultaneously for each headline. This effect may have had a stronger influence on discrimination for definitely true stimuli for the same reasons I described in the previous paragraph. What does appear apparent from my stimuli and design is that the truth judgments are stranger, or more challenging, than false judgments.

The stimuli in this study were certainly challenging to discriminate, perhaps because they were not always emotionally or politically charged or would have been challenging to quickly assess. It is also possible that some aspects of the stimuli used in this study may have made participants confused in a way that made them make more errors for specific stimulus classes. For example, the headline images, or who made an online statement, may have valenced participants in a way that took them away from assessing the content of the statement itself. Further, the stimulus class labels used in this study were changed from Politifact.com's original nominal labels to match the ordinal confidence scale to make them psychologically relevant. This might explain why participants showed a decrease for making definitely true confidence judgments for definitely true stimuli that Politifact.com simply labels "True"[7]. This could weaken the validity of my findings and may necessitate running the same study with different stimuli, at least in some stimulus classes. My goal was not to develop a psychometrically validated assessment most predictive of participants' ability to discriminate true and fake headlines because that has already been done (Maertens et al., 2023). However, this remains a great avenue of research to explore using the larger stimuli set from the norming study to select headlines to include in a future experiment based on how difficult headlines are and how well the headlines differentiate between participants' discrimination ability. Further assessing the psychometric properties of fact-checked headlines may provide deeper insights into what headlines, and components of headlines, are most predictive of an individual's ability to discriminate in the real world.

I found a significant interaction of order condition and stimulus class on confidence judgments showing that participants in the F-T condition often judged true headlines as more fake and fake headlines as more true relative to other conditions. This suggests that order can enhance the effect of headline veracity on participants' confidence judgments that are not present when examining order alone. This is an important finding that suggests misinformation could be designed in a particular pattern that spreads or reinforces falsehoods as facts or facts as falsehoods to gain viewer support. Participants may be conditioned to reject new information or cues when holding a belief. For instance, research has found that providing corrections for

[7] Politifact.com uses these labels: True, Mostly True, Half True, Mostly False, False, and Pants on Fire

misinformation may reinforce or anchor their beliefs in the original statement. Researchers call this phenomenon the backfire effect (Swire-Thompson et al., 2020). Depending on the source, participants may update their beliefs based on inaccurate statements. My study found some support for this claim by manipulating simple features of the headlines and the order in which headlines were presented.

In addition, I found that order did not influence participants' response bias. This was surprising because I developed *a priori* hypotheses that the T-F condition would induce a truth bias resulting in more willingness to make a "true" response to headlines despite their veracity, the F-T condition would induce a deception bias resulting in less willingness to make a "true" response to headlines, and the Random condition would be somewhere in the middle. A potential explanation for why I did not find any effect of order on response bias may be attributed to stable response tendencies that remain consistent regardless of the stimulus order. For example, an individual may have a general tendency, by their own nature, to be conservative or liberal in their willingness to make a particular response. Furthermore, participants could have rapidly adapted to the task to apply consistent response biases, canceling out any order effects. The stability of response bias has been studied by memory researchers and suggests bias is a cognitive trait that may be better explained by individual differences (see Kantner & Lindsay, 2012). From this research it appears that order may have a differential impact on cognitive processes involved in discrimination more than it affects decision-making strategies that are reflected in response bias.

Order did influence participants discrimination ability wherein the F-T condition resulted in lower discrimination relative to other conditions. This suggests that the F-T condition had an adverse effect on how participants diagnostically assess headline features. One explanation for this could be that the F-T condition impacted how participants process diagnostic features of the headlines. Using an eyewitness example, if police make a lineup with one black male and 5 white males, a witness may not carefully weigh the diagnostic features they remember from the perpetrator (e.g., an odd nose, a voice, a tattoo) and instead base their judgment on a non-diagnostic feature (e.g., "The guy that robbed me was black. This guy in the lineup is black, this must be him"). For instance, being initially exposed to definitely fake headlines (F-T condition) might induce weaker encoding of diagnostic features, or encoding of nondiagnostic features, that harm judgments on subsequent headlines. Conversely, initial exposure to definitely true headlines (T-F condition) may have induced participants to process stimuli deeper and encode a stronger representation of true stimuli. This may have improved participants' attention to more diagnostic features. This explanation is akin to a classical *serial-position effect* in cognitive psychology known as the *primacy effect* (see Murdock, 1962). For instance, headlines presented first are deeper encoded in participants' memory, serving as a comparison that may have harmed discrimination for participants in the F-T condition and improved discrimination for participants in the T-F condition. These findings suggest order can be a relevant factor for discriminating true and fake news and may be especially important considering the potential impacts misinformation has on real world decisions like who to vote for in an election or whether to get your children vaccinated.

A final interesting finding is the influence that individual differences have on all aspects of performance for discriminating true and fake headlines. Participants with greater belief in conspiracies, or reporting being more conservative consistently, across both studies, rated stimuli as more true on average, had poorer discrimination, and a greater willingness to make true responses regardless of the headline veracity relative to participants with lesser belief in conspiracies and reporting being more liberal. This finding supports much of the misinformation research that has focused on these individual differences (see Enders et al., 2021, 2023). I also found that other individual difference measures, often associated with belief in conspiracies and political orientation in the literature, also predict aspects of participants' performance including analytical thinking, information literacy, numeracy, open mindedness, and news knowledge.

Previous research suggests that individual differences may play a larger role in performance for true and fake news than previously realized and my findings support the same notion. For example, I found that conspiracy beliefs and political orientation explain differences in participants' confidence ratings. Participants with greater belief in conspiracies and reporting being more conservative showed a truth bias (tendency to rate headlines as true on average) relative to participants with lesser conspiratorial beliefs and reporting being more liberal that showed a deception bias (tendency to rate headlines as fake on average). This could be explained by research showing conservatives are more motivated to make fast and frugal judgments while liberals consider alternative response options in their confidence judgments (Ruisch & Stern, 2021). I also found that participant's information literacy scores explain their confidence ratings, but only in study 1 when source was not included. Interestingly, participants with higher information literacy scores showed a truth bias similar to beliefs in conspiracies and political orientation. This suggests that participants might rate headlines as true more often when they are conservative, believe in conspiracies, and are more skilled in understanding and processing information. Research exploring the utility of confidence in recognition memory accuracy suggests that individual differences account for a proportion of variance in confidence ratings equal to (for old judgments) or much larger (for new judgments) than response accuracy. This suggests that who makes the judgment is more important than accuracy alone, especially when making a "new" judgment (Kanter & Dobbins, 2019). This research may translate to the true versus fake judgments in a similar way. For instance, individual differences in political orientation might explain a larger proportion of variance in confidence ratings for fake judgments than true judgments.

I found that conspiracy belief and political orientation also explained differences in discrimination. Discrimination was worse for participants with greater conspiracy beliefs and reporting being more conservative. This could be explained by the same underpinnings that drive confidence judgments. Participants who are more conservative than liberal use different decision making strategies to rate confidence, which ultimately results in lower discrimination. Participants who have greater belief in conspiracies may also be reinforced by political ideology or, in some cases, may not have the cognitive resources to analytically defend themselves against misinformation. For study 1, analytical thinking had a positive effect on discrimination, which is consistent with the majority of the literature (see Pennycook & Rand, 2019). This supports a cognitive account of misinformation susceptibility. However, in study 2, I found a negative

relationship between analytical thinking and discrimination. This supports a motivational account of misinformation susceptibility (Pennycook & Rand, 2020). Overall, my findings support previous research suggesting that the psychological accounts of factors making people susceptible to misinformation are not mutually exclusive.

I found that when source was included, other factors such as numeracy skills, information literacy, and active open mindedness were significantly related to participants' discrimination ability. These findings fit the literature regarding how these measures should be related to discrimination. For example, participants that are more statistically numerate (and exhibiting more cognitive reflection) should have more efficient cognitive resources to evaluate the probability of headlines being true or fake resulting in better discrimination than those that have less numeracy skills. However, in study 2, I found a positive relationship between numeracy and discrimination but a negative relationship between analytical thinking and discrimination. Following the classical account of misinformation susceptibility, both analytical thinking and numeracy skills should both have a positive relationship with discrimination, and if they do not, it supports an integrated or motivational account of misinformation susceptibility. My results, particularly when source was included, appear to support a motivational account, which suggests misinformation susceptibility is driven by not only cognitive factors but also social identity factors such as political orientation and myside bias (see Roozenbeek et al., 2020). For instance, research finds that people view others, but not themselves, as biased (Wang & Jeon, 2020). This is particularly interesting considering that research shows that the integrated motivational account is supported when using various response modes and question framing (Roozenbeek et al., 2022). This switch could be due to the headline source engaging more cognitive resources to evaluate the credibility of the headline. When source was not included, participants may have needed to rely on their analytical ability more but rely on source as a heuristic cue when it is included. My results, in combination with the previous research, support the psychological explanation that misinformation susceptibility is not purely driven by cognitive factors but also by social factors that motivate individuals to process information in different ways, which has evident influences on aspects of performance.

Finally, response bias was explained by belief in conspiracies and political orientation in both studies. Participants with greater belief in conspiracies and reporting being more conservative needed less evidence to make a true response on average relative to participants with less belief in conspiracies and reporting being more liberal. Only in study 1, I found a significant negative relationship for information literacy on response bias again. However, I also found a positive relationship between social media usage and response bias. Participants who used social media more frequently required more evidence to make true judgments. This suggests that, on average, participants who use social media platforms more frequently are more cautious when responding true to headlines. This might be explained by the fact that frequent social media users are repeatedly exposed to high volumes of misinformation and fake news that make them more skeptical of, or have heightened vigilance for, headlines, which result in a more conservative response bias. Only in study 2, response bias was also explained by participants' news knowledge. Participants with greater knowledge about local, national, and world news needed less evidence to respond true to headlines regardless of their veracity. The positive

relationship between social media usage on response bias in study 1 and the negative relationship of news knowledge in study 2 is interesting because it suggests that being a frequent social media user may make you more cautious to make a true judgment while having more news knowledge makes you more likely to make a true judgment. This might be explained by the fact that social media users have more distrust for information they find on platforms, which induces a conservative response bias, whereas participants who rate they have more knowledge about local, national, and world news may receive that information from reliable sources, which increases their trust and tendency to respond true to headlines. Research showing that belief in fake news may be driven by an individual's tendency to be accepting of weak claims supports this explanation (Pennycook & Rand, 2020). My findings on how individual differences explain aspects of performance for detecting fake news show a consistent influence of belief in weak claims and political orientation on all aspects of performance. This finding, in combination with the others individual differences that influence performance across my studies, further support an integrated motivational account of misinformation susceptibility and that the absence/presence of headline source appears to alter what individual differences influence aspects of performance.

Overall, I found that participants can discriminate between fact-checked true and fake headlines, but with some caveats. Definitely true headlines confused participants and resulted in more errors and lower discrimination. Order was not important unless the degree to which fact-checked rating changed. Moreover, the F-T condition may influence participants' decision strategies, resulting in worse discrimination. I found that belief in conspiracies and political orientation consistently explained the proportion of variance in all aspects of performance. The other individual differences also made contributions explaining performance and provide support for the theoretical frameworks described in the literature. A summary of my hypothesis findings for both studies can be found in Table 11.

Table 11.

Summary of Hypothesis Findings from Study 1 and Study 2

Hypothesis	Study 1 Findings	Study 2 Findings
1	Found general support that discrimination differs across stimulus class with a decrease for definitely true headlines. Discrimination was lowest for participants in the fake-true order condition	Found general support that discrimination differs across stimulus class with a decrease for definitely true headlines. Discrimination was lowest for participants in the fake-true order condition
2	Cognitive reflection, conspiracy beliefs, and political orientation significantly predicted participants discrimination	Cognitive reflection, conspiracy beliefs, political orientation, numeracy skills, information literacy, and open-mindedness significantly predicted participants discrimination
3	No support that response bias differed by order	No support that response bias differed by order
4	Information literacy scores, social media usage, conspiracy beliefs, and political orientation significantly predicted participants response bias	News knowledge, conspiracy beliefs, and political orientation significantly predicted participants response bias.
5	No support that calibration was better for true than fake headlines. Calibration was better for fake than true headlines	No support that calibration was better for true than fake headlines. Calibration was better for fake than true headlines
6	Information literacy scores, conspiracy beliefs and political orientation significantly predicted participants confidence judgments	Conspiracy beliefs and political orientation significantly predicted participants confidence judgments

Strengths and Limitations

I randomly assigned participants to order conditions and manipulated the headline veracity from fact-checked headlines to assess performance differences. The methodology I used in my studies is a particular strength of my findings. While this study replicated and extended most of my findings from study 1, by including the original headline source in study 2, my studies have limitations that should be addressed before accepting my findings as reliable and valid.

To my knowledge, this study is the first of its kind utilizing headlines from a fact-checking website, manipulating the degree of truthfulness or falsity for headlines, and manipulating the order in which headlines are presented to participants. I used true and fake headlines from Politifact.com – a fact-checking website. The 30 true and fake headlines used in the studies were headlines that came from a larger set of 90 true and fake headlines. Headlines

were subjected to a norming study and headlines that participant's ratings reasonably matched participant's confidence were included in both studies. Interestingly, ChatGPT showed similar results. I believe this is an advantage of my study because headlines were quantitatively evaluated before data collection and the headlines contained statements made by public officials or organizations in the real world while previous research often included stimuli that are made up or only includes text. In addition, including fact-checked headlines in my study reinforces findings from the literature because my results reflect headlines that participants can, or even may have, encountered in the real world.

As mentioned in the introduction, previous research has primarily investigated headlines that are considered absolutely true or fake but do not explore headlines that are graded in the degree of truthfulness or falsity. In the real world, statements are unlikely to be entirely fact or fiction but instead could contain traces or aspects of truth and falsehoods. The benefits of including headlines that are graded in their veracity strengthens my findings because it reflects real-world information and provides a more nuanced appreciation of how participants process and respond to true and fake headlines. My findings suggest that using headlines that vary in the degree of veracity contribute to understanding how participants are influenced by the ambiguity or certainty of headline's veracity.

I found consistent findings across both studies. However, the six headline stimulus classes had only five headlines each, which could have impacted the validity of its effects on performance. I reduced the total number of headlines used in each stimulus class in the norming study but may have benefited if I had more qualifying stimuli for each class. As mentioned, I changed the Likert labels Politifact.com gave to their fact-checked headlines to make them psychologically relevant. This change may have influenced how participants rated their confidence because Politifact.com's rating system provides a much higher degree of certainty for fake headlines (Politifact.com uses "Pants on Fire" as the highest rating) than true headlines ("True" is the highest rating). If Politifact.com's rating are accurate, that could contribute to the odd findings associated with the definitely true stimuli. In addition, I may have observed better discrimination if I had included a neutral option. Forcing participants to make confidence judgments that fall into a binary group (true or fake) may have reduced discrimination overall. But to reiterate, the purpose of this study was to evaluate how well participants could discriminate true and fake news, not to criticize or evaluate how accurately Politifact.com labeled or fact-checked headlines.

In addition, I investigated how stimulus order influences aspects of performance for true and fake headlines. As described, order serves as a proxy for task difficulty that introduces cognitive load on participants. Stimulus order is a classic, but often overlooked, manipulation that can be insightful into how adding complexity to a task influences decision making processes. I found that order can significantly change across stimulus class and influences confidence judgments when order interacts with headline veracity. These findings make contributions to just how challenging discriminating true and fake information might be in the real world. Subtle stimulus effects, like changes in the order headlines are presented to participants, and how fact-checked headlines vary in veracity, can have differential influences on

participants' confidence judgments, their tendency to make a true/fake response, and their ability to discriminate true from fake headlines.

Order is important because bad actors could manipulate order if the aim was to hide or dilute the truth. In relation to misinformation, order mimics scenarios where individuals may be initially exposed to highly certain fake headlines, highly certain true headlines, or may be randomly exposed to true and fake headlines. I could have manipulated both order and the base rate of true and fake headlines. This would have likely shown influences on participants' response bias. Insights from my research could be expanded into a recognition memory framework by having participants study headlines in different order conditions and allow them to make confidence judgments for new and old true and fake headlines at differing base rates of fake headlines. Doing so would lend itself to deeper understanding of how influential order is and how memory processes contribute to the misinformation context. Moreover, doing so would allow for more complex modeling techniques that could be useful for understanding participants memory and decision making processes in making true and fake judgments.

It is also possible that participating in a study could be an explanation for some of my findings. For instance, participants were aware they were in a study, which may have discouraged them to perform in a genuine manner, which may have hampered my ability to assess my hypotheses. As mentioned, participants could have judged their confidence using different decision making strategies. Research shows that various stimulus and task effects can influence whether participants treat memory evidence continuously or discretely (McAdoo et al., 2019). For example, some participants may make true and fake judgments in different ways and prefer to rate confidence with high certainty using the extreme values on the scale, while other participants may make ratings with low certainty using the middle values on the scale. If I had designed my study in a way that manipulated the amount of exposure participants had to true and fake news to vary their memory strength for studied headlines, I would have been able to model whether participants in groups discriminated true and fake headlines discretely (all or none) or continuously. This would be important to examine in future research and could be theoretically relevant to better understanding recognition memory judgments.

Aspects of my research design may have also negatively influenced participants' performance. Presenting participants with both sharing and accuracy items together was done to adversely impact accuracy because considering sharing reduces an individual's focus on accuracy (see Epstein et al., 2021). This methodology could have influenced whether participants processed headlines by an automatic or controlled process but I felt that it was important to include because it reflects how the social media context interferes with discrimination. I also chose to use only a single measure of confidence with a 1-6 Likert scale as to not fatigue participants and the judgments could be easily transformed into binary judgments (ratings of 1, 2, or 3 correspond to fake judgments and ratings of 4, 5, or 6 correspond to true judgments). I might have found different results if I designed the study to have participants first make binary decisions about whether the headline was true or fake followed by a 0-100 confidence judgment. Nonetheless, I did find similar results across two experiments, which supports my findings, but this could have been due to the methodology I used. Overall, I used a methodology informed by

the previous literature and used quantitative models to evaluate how stimulus factors and individual differences explain aspects of performance.

Future Research

The methodology used in this study is based on and extends the theoretical framework described in the literature. However, additional research should be done to overcome the limitations of the current research findings and take various considerations into account.

Researchers can take my findings and choose which direction to take next, but I recommend doing further research that gets at how people are realistically interacting with news headlines. I could have utilized a different approach to selecting stimuli from my norming study. I would have liked to had stimuli that varied in how congruent/incongruent it was with an individual's ideology, but these were hard to find without making them myself. By using pre-existing stimuli, I was not able to manipulate how Democratic/Republican a headline was. A stimuli set that does include this factor would be very valuable because it would allow researchers to attempt to replicate and extend my findings. Additional research efforts should be made to develop larger and current databases of real-world true and fake stimuli that include other features related to the headline (e.g., source, political association, valence) so the psychometric properties of headlines can be assessed. For example, item analysis, a psychometric technique, would allow me to answer questions like, "which headlines were more predictive of an individual's latent ability to discriminate true and fake news?". These stimuli should be normed and graded on multiple dimensions. Doing so would allow researchers to use a common database to investigate the influence of various factors and conduct more complex modeling techniques on the stimuli.

Misinformation research has exploded across multiple disciplines, especially following the onset of COVID-19. If misinformation is as important as is communicated to the public, then scientists must carefully examine the methods used to evaluate its influence. Psychological phenomenon, like discriminating true and fake news, are often explained using verbal theories, which cannot be substituted for quantitative analysis. Cognition and behavior are probabilistic and have underlying principles that can be better understood through formalization (e.g., mathematical equations and rules), because patterns in data have causes (Farrell & Lewandowsky, 2018; Ward, 2002). The problem is, almost any data can be interpreted post-hoc to fit verbal theories, making them difficult to falsify, which slows the field's scientific progress.

I made efforts to systematically control factors I was interested in and evaluated my data using analyses commonly found in misinformation research like mixed effects regression to control for Type I error rates and, to a lesser extent, signal detection. But my decision in doing what I did was to account for differences in performance outcomes between participants to reduce Type I error rates; other regression models and techniques may have found different results. For example, I could have performed structural equation modeling (SEM), but my main goal was not to confirm the causal relationship between my variables. Rather, my goal was to explore what predictors explained individuals' confidence judgments and objective performance outcomes.

Data from research should be made easily available to replicate findings and use advanced modeling techniques. For instance, Ratcliff and Starns (2009) developed a diffusion model (RTCON) for confidence judgments in recognition memory. In the model, evidence from a stimulus is accumulated to decision criteria that corresponds to different confidence levels. However, the RTCON model, and other diffusion models, require response time data. Collecting and analyzing response time data in combination with confidence ratings could be very illuminating in the misinformation domain.

From my literature search, SDT has not been used much in misinformation research, which was surprising because it is a common technique used to detect a signal among noise across various domains including experimental psychology, medicine, and military (see Wixted, 2020). Although SDT models are a common technique that have been applied to various domains, the basic SDT model assumes the signal and noise distributions are normal distributions and, typically, have equal variances. This assumption places limitations on the generalizability of the model because it does not accurately reflect many real-world problems. For example, empirical evidence has demonstrated that variance in the signal distribution is larger than in the noise distribution (Ratcliff, Sheu, & Gronlund, 1992). This issue has presented some criticism to using the model as a reliable technique for modeling recognition performance.

In recognition performance, one approach to obtaining more informative data is having participants judge their confidence of correctly identifying each item on a Likert scale (e.g., "how confident are you that the word presented was studied?") instead of or in addition to the simple categorical decision ("was the word studied or not?"). However, the issue with confidence judgments is that the number of thresholds that have to be estimated increases with the number of Likert scale categories. For example, if the model includes confidence data from a seven-point Likert scale, then eight parameters – *d'*, σ, and six thresholds (*c*) – have to be estimated. Moreover, the reliable estimation of additional thresholds requires large datasets with a minimum number of observations for each condition, meaning that models with more Likert scale categories require larger datasets (Selker et al., 2019), which can result in fatigue effects for participants across trials or conditions (Criss et al., 2011). Further, some participants prefer to use the outer values of the scale, while other prefer to use the inner values. To account for these individual differences, a model needs to estimate the thresholds set by a participant to choose a certain response.

Selker et al. (2019) described a parsimonious method for estimating confidence thresholds by restricting how the thresholds can be placed. This parsimony is obtained by modeling the thresholds as linear transformations of unbiased thresholds, which only requires two parameters for any number of thresholds. The scale parameter (*a*) allows the thresholds to be distributed more closely to the center of the scale or further away. Shift (*b*) allows the thresholds to focus more on the left or right side of the scale. Changes in shift and scale can account for various ways thresholds are placed and can extend to any number of thresholds without requiring additional parameters. This is important because stronger models allow for deeper insights into how people make decisions when discriminating true and fake headlines.

I believe another important next step is to determine how stimulus order, stimulus class, and individual differences influence the effectiveness of interventions. Many research groups across multiple disciplines have examined treatment effectiveness but would benefit from producing more ecologically valid assessments of the adverse effects of misinformation. If researchers aim to gain a deeper understanding into how well people discriminate true and fake news in the real world, then an individual's discrimination ability should be intensively sampled across time at random. Ecological momentary assessment (EMA) is often used in mobile health research and may be a beneficial approach to assessing misinformation discrimination. This approach, using psychometrically validated headlines and robust quantitative methods, would provide more ecologically valid results when assessing interventions to improve discrimination or reduce susceptibility to misinformation.

Conclusions

On a daily basis, people often engage with multiple social media platforms and are exposed to high volumes of information which, naturally, increases chances of encountering misinformation. The main goal of the current research was to evaluate participants ability to discriminate true from fake headlines, the influence of stimulus order on participants confidence judgments, and what individual differences were good predictors of performance. The findings highlight that participants were mediocre at discriminating true from fake headlines overall but was likely due to reduced discrimination for definitely true headlines. Participants were better calibrated for fake headlines than true headlines. The effect of stimulus order on confidence judgments was dependent on the stimulus class. Participants' beliefs in conspiracies, political orientation, and other individual differences explained variance in all aspects of performance. Overall, I found consistent performance results across two studies, without and with headline source included. My studies provide nuanced insights into participants ability to discriminate true from fake news headlines and potential explanations for participants decision making. I also provide methodological considerations for future research.

References

Alimohamadi, Y., & Sepandi, M. (2019). Considering the design effect in cluster sampling. *Journal of Cardiovascular and Thoracic Research*, *11*(1), 78-78.

Altay, S., Lyons, B., & Modirrousta-Galian, A. (2023). Exposure to higher rates of false news erodes media trust and fuels skepticism in news judgment. *PsyArXiv*.

Bago, B., Rand, D. G., & Pennycook, G. (2022). Does deliberation decrease belief in conspiracies?. *Journal of Experimental Social Psychology*, *103*, 104395.

Barajas, J., & John, R. (2023). A Signal Detection Theory Approach to Predicting Immunity to Pandemic Vaccine Fake News. In *Proceedings of the 54th Hawaii International Conference on System Sciences*.

Baron, J., Isler, O., & Yilmaz, O. (2022). Actively open-minded thinking and the political effects of its absence. *Psyarxiv*.

Basol, M., Roozenbeek, J., & Van der Linden, S. (2020). Good news about bad news: Gamified inoculation boosts confidence and cognitive immunity against fake news. *Journal of Cognition, 3*(1).

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.

Benjamin, A. S. (2005). Recognition memory and introspective remember/know judgments: Evidence for the influence of distractor plausibility on "remembering" and a caution about purportedly nonparametric measures. *Memory & Cognition, 33*, 261-269.

Benjamin, A. S., & Bawa, S. (2004). Distractor plausibility and criterion placement in recognition. *Journal of Memory and Language, 51*(2), 159-172.

Boh Podgornik, B., Dolničar, D., Šorgo, A., & Bartol, T. (2016). Development, testing, and validation of an information literacy test (ILT) for higher education. *Journal of the Association for Information Science and Technology*, *67*(10), 2420-2436.

Brotherton, R., French, C. C., & Pickering, A. D. (2013). Measuring belief in conspiracy theories: The generic conspiracist beliefs scale. *Frontiers in Psychology, 4*, 279.

Brown, S., & Steyvers, M. (2005). The dynamics of experimentally induced criterion shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(4), 587.

Byrne, A., Martin, D., Jones, C., Galbraith, N., & Mercer, T. (2024). Conspiracy theory beliefs in the adolescent population: A systematic review. *Journal of Adolescence*.

Calvillo, D. P., Garcia, R. J., Bertrand, K., & Mayers, T. A. (2021). Personality factors and self-reported political news consumption predict susceptibility to political fake news. *Personality and Individual Differences, 174*, 110666.

Clayton, K., Blair, S., Busam, J. A., Forstner, S., Glance, J., Green, G., ... & Sandhu, M. (2020). Real solutions for fake news? Measuring the effectiveness of general warnings and factcheck tags in reducing belief in false stories on social media. *Political Behavior, 42*, 1073-1095.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin numeracy test. *Judgment and Decision Making, 7*(1), 25-47.

Colloff, M. F., Wade, K. A., & Strange, D. (2016). Unfair lineups make witnesses more likely to confuse innocent and guilty suspects. *Psychological Science, 27*(9), 1227-1239.

Criss, A. H., Malmberg, K. J., & Shiffrin, R. M. (2011). Output interference in recognition memory. *Journal of Memory and Language, 64*(4), 316–326.

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences, 113*(3), 554-559.

Dougal, S., & Rotello, C. M. (2007). "Remembering" emotional words is based on response bias, not recollection. *Psychonomic Bulletin & Review, 14*(3), 423-429.

Ecker, U., Roozenbeek, J., van der Linden, S., Tay, L. Q., Cook, J., Oreskes, N., & Lewandowsky, S. (2024). Misinformation poses a bigger threat to democracy than you might think. *Nature, 630*(8015), 29-32.

Enders, A., Farhart, C., Miller, J., Uscinski, J., Saunders, K., & Drochon, H. (2023). Are republicans and conservatives more likely to believe conspiracy theories?. *Political Behavior, 45*(4), 2001-2024.

Enders, A. M., Uscinski, J. E., Seelig, M. I., Klofstad, C. A., Wuchty, S., Funchion, J. R., ... & Stoler, J. (2021). The relationship between social media use and beliefs in conspiracy theories and misinformation. *Political Behavior*, 1-24.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: a new look at an old issue. *Psychological Methods, 12*(2), 121.

Epstein Z., Sirlin N., Arechar A. A., Pennycook G., Rand D. G. (2021). The Social Media Context Interferes with Truth Discernment. *Science Advances, 9*(9).

Farrell, S., & Lewandowsky, S. (2018). *Computational Modeling of Cognition and Behavior*. Cambridge University Press.

Fischer, H., Amelung, D., & Said, N. (2019). The accuracy of German citizens' confidence in their climate change knowledge. *Nature Climate Change, 9*(10), 776-780.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25-42.

Garson, G. D. (2013). *Hierarchical linear modeling: Guide and applications*. SAGE Publications.

Gawronski, B., Ng, N. L., & Luke, D. M. (2023). Truth sensitivity and partisan bias in responses to misinformation. *Journal of Experimental Psychology: General*.

Gronlund, S. D., & Benjamin, A. S. (2018). The new science of eyewitness memory. In *Psychology of Learning and Motivation* (Vol. 69, pp. 241-284). Academic Press.

Heit, E., Brockdorff, N., & Lamberts, K. (2003). Adaptive changes of response criterion in recognition memory. *Psychonomic Bulletin & Review, 10*(3), 718-723.

Helm, R. K., & Nasu, H. (2021). Regulatory responses to 'fake news' and freedom of expression: Normative and empirical evaluation. *Human Rights Law Review, 21*(2), 302-328.

Jones-Jang, S. M., Mortensen, T., & Liu, J. (2021). Does media literacy help identification of fake news? Information literacy helps, but other literacies don't. *American Behavioral Scientist, 65*(2), 371-388.

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy, 1*(1), 54-86.

Kantner, J., & Dobbins, I. G. (2019). Partitioning the sources of recognition confidence: The role of individual differences. *Psychonomic Bulletin & Review, 26*, 1317-1324.

Kantner, J., & Lindsay, D. S. (2012). Response bias in recognition memory as a cognitive trait. *Memory & Cognition, 40*, 1163–1177.

Kay, C. S., & Slovic, P. (2023). The generic conspiracist beliefs scale–5: A short-form measure of conspiracist ideation. *Journal of Research in Personality, 102*, 104315.

Kemp, P. L., Alexander, T. R., & Wahlheim, C. N. (2022). Recalling fake news during real news corrections can impair or enhance memory updating: the role of recollection-based retrieval. *Cognitive Research: Principles and Implications, 7*(1), 1-21.

Layher, E., Dixit, A., & Miller, M. B. (2020). Who gives a criterion shift? A uniquely individualistic cognitive trait. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(11), 2075.

Lazer, D. M., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., ... & Zittrain, J. L. (2018). The science of fake news. *Science, 359*(6380), 1094-1096.

Lorenz-Spreen, P., Lewandowsky, S., Sunstein, C. R., & Hertwig, R. (2020). How behavioural sciences can promote truth, autonomy and democratic discourse online. *Nature human Behaviour, 4*(11), 1102-1109.

Luo, M., Hancock, J. T., & Markowitz, D. M. (2022). Credibility perceptions and detection accuracy of fake news headlines on social media: Effects of truth-bias and endorsement cues. *Communication Research, 49*(2), 171-195.

Macmillan, N. A., & Creelman, C. D. (2004). *Detection Theory: A User's Guide*. Psychology press.

Maertens, R., Götz, F. M., Schneider, C. R., Roozenbeek, J., Kerr, J. R., Stieger, S., McClanahan, W. P., Drabot, K., & van der Linden, S. (2022). The Misinformation Susceptibility Test (MIST): A psychometrically validated measure of news veracity discernment. *PsyArXiv*.

McAdoo, R. M., Key, K. N., & Gronlund, S. D. (2019). Task effects determine whether recognition memory is mediated discretely or continuously. *Memory & Cognition, 47*, 683-695.

McNeish, D. M., & Stapleton, L. M. (2016). The effect of small sample size on two-level model estimates: A review and illustration. *Educational Psychology Review, 28*, 295-314.

Michal, A. L., Zhong, Y., & Shah, P. (2021). When and why do people act on flawed science? Effects of anecdotes and prior beliefs on evidence-based decision-making. *Cognitive Research: Principles and Implications, 6*(1), 1-23.

Mickes, L., Wixted, J. T., & Wais, P. E. (2007). A direct test of the unequal-variance signal detection model of recognition memory. *Psychonomic Bulletin & Review, 14*(5), 858-865.

Miller, T. M., & Geraci, L. (2011). Training metacognition in the classroom: The influence of incentives and feedback on exam predictions. *Metacognition and Learning, 6*, 303-314.

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*.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology, 64*(5), 482-488.

Pantazi, M., Kissine, M., & Klein, O. (2018). The power of the truth bias: False information affects memory and judgment even in the absence of distraction. *Social Cognition, 36*(2), 167-198.

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*, 1-12.

Pennycook, G., Bago, B., & McPhetres, J. (2022). Science beliefs, political ideology, and cognitive sophistication. *Journal of Experimental Psychology: General*.

Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General, 147*(12), 1865.

Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2020a). On the belief that beliefs should change according to evidence: Implications for conspiratorial, moral, paranormal, political, religious, and science beliefs. *Judgment and Decision Making, 15*(4), 476-498.

Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020b). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological Science, 31*(7), 770-780.

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.

Pennycook, G., & Rand, D. G. (2020). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality, 88*(2), 185–200.

Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences, 25*(5), 388-402.

Pew Research Center. (2020, April 20). From virtual parties to ordering food, how Americans are using the internet during COVID-19 [Report]. https://www.pewresearch.org/short-reads/2020/04/30/from-virtual-parties-to-ordering-food-how-americans-are-using-the-internet-during-covid-19/.

Rapp, D. N., & Withall, M. M. (2023). Confidence as a metacognitive contributor to and consequence of misinformation experiences. *Current Opinion in Psychology*, 101735.

Ratcliff, R., Sheu, C. F., & Gronlund, S. D. (1992). Testing global memory models using ROC curves. *Psychological Review, 99*(3), 518.

Ratcliff, R., & Starns, J. J. (2009). Modeling confidence and response time in recognition memory. *Psychological Review, 116*(1), 59.

Roozenbeek, J., Maertens, R., Herzog, S. M., Geers, M., Kurvers, R., Sultan, M., & van der Linden, S. (2022). Susceptibility to misinformation is consistent across question framings and response modes and better explained by myside bias and partisanship than analytical thinking. *Judgment and Decision Making, 17*(3), 547-573.

Roozenbeek, J., Schneider, C. R., Dryhurst, S., Kerr, J., Freeman, A. L., Recchia, G., van der Bles, A. M., & Van Der Linden, S. (2020). Susceptibility to misinformation about COVID-19 around the world. *Royal Society Open Science, 7*(10), 201199.

Ross, B., Jung, A., Heisel, J., & Stieglitz, S. (2018). Fake news on social media: The (in) effectiveness of warning messages. *Thirty Ninth International Conference on Information Systems, San Francisco 2018*. Association for Information Systems.

Rotello, C. M., & Macmillan, N. A. (2007). Response bias in recognition memory. *Psychology of Learning and Motivation, 48*, 61-94.

Ruisch, B. C., & Stern, C. (2021). The confident conservative: Ideological differences in judgment and decision-making confidence. *Journal of Experimental Psychology: General, 150*(3), 527.

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.

Schwartz, L. M., Woloshin, S., Black, W. C., & Welch, H. G. (1997). The role of numeracy in understanding the benefit of screening mammography. *Annals of Internal Medicine, 127*(11), 966-972.

Selker, R., van den Bergh, D., Criss, A. H., & Wagenmakers, E. J. (2019). Parsimonious estimation of signal detection models from confidence ratings. *Behavior Research Methods, 51*, 1953-1967.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers, 31*(1), 137-149.

Stretch, V., & Wixted, J. T. (1998). Decision rules for recognition memory confidence judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*, 1397–1410.

Sultan, M., Tump, A. N., Geers, M., Lorenz-Spreen, P., Herzog, S. M., & Kurvers, R. H. (2022). Time pressure reduces misinformation discrimination ability but does not alter response bias. *Scientific Reports, 12*(1), 22416.

Swets, J. A., Dawes, R. M., & Monahan, J. (2000). Psychological science can improve diagnostic decisions. *Psychological Science in the Public Interest, 1*(1), 1-26.

Swire-Thompson, B., Ecker, U. K., Lewandowsky, S., & Berinsky, A. J. (2020). They might be a liar but they're my liar: Source evaluation and the prevalence of misinformation. *Political Psychology, 41*(1), 21-34.

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making, 11*(1), 99–113.

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.

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.

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.

Van der Linden, S., Panagopoulos, C., Azevedo, F., & Jost, J. T. (2021). The paranoid style in American politics revisited: An ideological asymmetry in conspiratorial thinking. *Political Psychology, 42*(1), 23-51.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146-1151.

Wang, Q., & Jeon, H. J. (2020). Bias in bias recognition: People view others but not themselves as biased by preexisting beliefs and social stigmas. *PLoS One, 15*(10), e0240232

Walter, N., & Tukachinsky, R. (2020). A meta-analytic examination of the continued influence of misinformation in the face of correction: How powerful is it, why does it happen, and how to stop it?. *Communication Research, 47*(2), 155-177.

Ward, L. M. (2002). *Dynamical Cognitive Science*. MIT press.

Wixted, J. T. (2020). The forgotten history of signal detection theory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 46*(2), 201.

Wixted, J. T., Mickes, L., Wetmore, S. A., Gronlund, S. D., & Neuschatz, J. S. (2017). ROC analysis in theory and practice. *Journal of Applied Research in Memory and Cognition, 6*(3), 343-351.

Wixted, J. T., & Wells, G. L. (2017). The relationship between eyewitness confidence and identification accuracy: A new synthesis. *Psychological Science in the Public Interest, 18*(1), 10-65.

Appendix

Study Instructions

Subjects were presented with the following instructions:

"The purpose of this study is to assess peoples' perceptions of news headlines online. There are no tricks involved and there is a 50/50 chance any given headline you judge will be true or fake. You will be presented with a series of news headlines seen online and perform the following for each headline: 1) your likelihood to share each news headline on social media and 2) rate how confident you are each news headline is true or fake without using human and non-human resources (including the internet). Press the next button when you are ready to begin the study."

MIST-8 Items (Maertens et al., 2022; Roozenbeek et al., 2022)

Rate how confident you are the following headline is true or fake (1 = Definitely Fake, 2 = Likely Fake, 3 = Probably Fake, 4 = Probably True, 5 = Likely True, 6 = Definitely True)

F1 Government Officials Have Manipulated Stock Prices to Hide Scandals

F3 New Study: Left-Wingers Are More Likely to Lie to Get a Higher Salary

F6 Certain Vaccines Are Loaded with Dangerous Chemicals and Toxins

F8 The Government Is Knowingly Spreading Disease Through the Airwaves and Food Supply

R1 Attitudes Toward EU Are Largely Positive, Both Within Europe and Outside It

R5 Hyatt Will Remove Small Bottles from Hotel Bathrooms by 2021

R7 Republicans Divided in Views of Trump's Conduct, Democrats Are Broadly Critical

R9 Global Warming Age Gap: Younger Americans Most Worried

CRT Items

Items from Frederick (2005)

(1) A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? _____ cents. (Correct response = 5 cents; Intuitive response = 10 cents)
(2) If it takes 5 machines 5 min to make 5 widgets, how long would it take 100 machines to make 100 widgets? _____ min. (Correct response = 5 min; Intuitive response = 100 min)
(3) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? _____ days. (Correct response = 47 days; Intuitive response = 24 days)

Items from Thomson & Oppenheimer (2016)

(4) If you're running a race and you pass the person in second place, what place are you in? (Correct response = second; Intuitive response = first)
(5) A farmer had 15 sheep and all but 8 died. How many are left? (Correct response = 8; Intuitive response = 7)
(6) Emily's father has three daughters. The first two are named April and May. What is the third daughter's name? (Correct response = Emily; Intuitive response = June)
(7) How many cubic feet of dirt are there in a hole that is 3' deep x 3' wide x 3' long? (Correct response = none; Intuitive response = 27)

Numeracy

Schwartz et al. (1997) items

(1) Imagine that we flip a fair coin 1,000 times. What is your best guess about how many times the coin would come up heads in 1,000 flips? _____ times out of 1,000. (Correct response = 500)
(2) In the BIG BUCKS LOTTERY, the chance of winning a $10 prize is 1%. What is your best guess about how many people would win a $10 prize if 1000 people each buy a single ticket to BIG BUCKS? _____ person(s) out of 1,000. (Correct response = 10)
(3) In ACME PUBLISHING SWEEPSTAKES, the chance of winning a car is 1 in 1,000. What percent of tickets to ACME PUBLISHING SWEEPSTAKES win a car? _____ % (Correct response = 0.1)

Items from Berlin numeracy test (Cokely et al., 2012)

(1) Imagine we are throwing a five-sided die 50 times. On average, out of these 50 throws how many times would this five-sided die show an odd number (1, 3 or 5)? _____ out of 50 throws. (Correct response = 30)
(2) Out of 1,000 people in a small town 500 are members of a choir. Out of these 500 members in the choir 100 are men. Out of the 500 inhabitants that are not in the choir 300 are men. What is the probability that a randomly drawn man is a member of the choir? (please indicate the probability in percent). _____ % (Correct response = 25)
(3) Imagine we are throwing a loaded die (6 sides). The probability that the die shows a 6 is twice as high as the probability of each of the other numbers. On average, out of these 70 throws, how many times would the die show the number 6? _____ out of 70 throws. (Correct response = 20)
(4) In a forest 20% of mushrooms are red, 50% brown and 30% white. A red mushroom is poisonous with a probability of 20%. A mushroom that is not red is poisonous with a probability of 5%. What is the probability that a poisonous mushroom in the forest is red? _____% (Correct response = 50)

Single item from Roozenbeek et al., 2020, 2022

(1) What represents the highest chance of something happening: 1 in 10, 1 in 1000, or 1 in 100. (Correct response = 1 in 10)

AOT (Baron et al., 2022)

1-5 scale: 1 = Completely disagree; 5 = Completely agree. (–) indicates reverse-scored items.

1. Willingness to be convinced by opposing arguments is a sign of good character.
2. People should take into consideration evidence that goes against conclusions they favor.
3. Being undecided or unsure is the result of muddled thinking. (–)
4. People should revise their conclusions in response to relevant new information.
5. Changing your mind is a sign of weakness. (–)
6. People should search actively for reasons why they might be wrong.
7. It is OK to ignore evidence against your established beliefs. (–)
8. It is important to be loyal to your beliefs even when evidence is brought to bear against them. (–)
9. There is nothing wrong with being undecided about many issues.
10. When faced with a puzzling question, we should try to consider more than one possible answer before reaching a conclusion.

ILT Items (Jones-Jang et al., 2021)

1. The most reliable, verified, concise and comprehensive description of an unknown specialized concept can be found in _____. ((a) daily newspaper, (b) bilingual dictionary, **(c) lexicon or encyclopedia**, (d) research article)
2. In which list have the information sources been correctly ordered from the least to the most formally established and verified? (**(a) blog, daily newspaper, scholarly journal, standard**, (b) blog, standard, daily newspaper, scholarly journal, (c) daily newspaper, blog, standard, scholarly journal, (d) standard, scholarly journal, blog, daily newspaper)
3. Which of the data listed below are "raw" unprocessed data? (**(a) share prices published at the end of a trading day**, (b) weather maps, (c) population growth data presented in tables, (d) population growth data presented diagrammatically (in graphs).
4. Compared with a search within the title and abstract, a full-text search database results in _____. ((a) the same number of hits, (b) a smaller number of hits, (c) this has no effect on the number of hits, **(d) a larger number of hits**)
5. Which statement on GMO is not the author's personal opinion? ((a) GMO will bring about a global food crisis, **(b) According to inventories, 15 new GMOs were registered in the EU in 2013**, (c) GMO experimentation should be banned, (d) Most GMO researchers have been paid off by large corporations, such as Monsanto)

GCB-5 items (Kay & Slovic, 2023)

1-5 scale: 1 = Strongly disagree; 5 = Strongly agree

1. The government permits or perpetrates acts of terrorism on its own soil, disguising its involvement.
2. Evidence of alien contact is being concealed from the public.
3. New and advanced technology which would harm current industry is being suppressed.
4. Certain significant events have been the result of the activity of a small group who secretly manipulate world events.
5. Experiments involving new drugs or technologies are routinely carried out on the public without their knowledge or consent.

Social media use items

Rate how often you use the following social media platforms (from 1 “never” to 5 “all the time”)

- 4chan,
- Discord,
- Facebook messenger,
- Facebook,
- GETTR,
- Instagram,
- Omegle
- Parlor,
- Reddit,
- Snapchat,
- TikTok,
- Truth Social,
- Twitch,
- Twitter,
- WhatsApp,
- YouTube

Study 1 ANOVA Table for Confidence Judgments

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)	
Condition	1.54	0.769	2	455	2.3279	0.09866	.
Stimulus Class	479.73	95.946	5	2320	290.4728	< 2.2e-16	***
CRT	0.05	0.048	1	455	0.1461	0.70244	
Numeracy	0.04	0.038	1	455	0.1156	0.734	
ILT	1.38	1.38	1	455	4.1769	0.04155	*
GCB	20.57	20.57	1	455	62.2747	2.25E-14	***
AOT	0.52	0.521	1	455	1.5784	0.20964	
Social Media Use	0.06	0.056	1	455	0.1702	0.68009	
Political Orientation	11.2	11.203	1	455	33.9154	1.09E-08	***
News Knowledge	0	0.001	1	455	0.0044	0.94724	
Familiarity	0.15	0.153	1	455	0.4643	0.49596	
Condition:Stimulus Class	14.51	1.451	10	2320	4.3916	3.92E-06	***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Study 1 ANOVA Table for Discrimination

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)	
CRT	1.197	1.197	1	2790.3	3.9135	0.048	*
Numeracy	0.11	0.11	1	2790.4	0.3579	0.54974	
ILT	0.969	0.969	1	2790.3	3.1685	0.07518	.
GCB	31.035	31.035	1	2791.8	101.4298	< 2e-16	***
AOT	44.49	44.49	1	2790.5	145.4026	< 2e-16	***
Social Media Use	0.698	0.698	1	2790.1	2.2805	0.13112	
Political Orientation	101.427	101.427	1	2790.1	331.4858	< 2e-16	***
News Knowledge	0.532	0.532	1	2790.1	1.7392	0.18735	
Familiarity	1.114	1.114	1	2790.4	3.6416	0.05646	.

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Study 1 ANOVA Table for Response Bias

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)	
CRT	0.142	0.142	1	2791.1	0.6773	0.41058	
Numeracy	0.035	0.035	1	2791.4	0.1663	0.68346	
ILT	7.327	7.327	1	2791.1	35.0052	3.69E-09	***
GCB	62.851	62.851	1	2784.7	300.2715	< 2.2e-16	***
AOT	0.028	0.028	1	2791.7	0.1328	0.71553	
Social Media Use	1.135	1.135	1	2790.5	5.422	0.01996	*
Political Orientation	27.13	27.13	1	2790.5	129.6128	< 2.2e-16	***
News Knowledge	0.043	0.043	1	2790.4	0.2064	0.64963	
Familiarity	0.097	0.097	1	2791.4	0.4618	0.49682	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Study 2 ANOVA Table for Confidence Judgments

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)	
Condition	0.43	0.216	2	457	0.6813	0.506459	
Stimulus Class	458.72	91.743	5	2330	288.9551	< 2.2e-16	***
CRT	0.07	0.073	1	457	0.2306	0.631336	
Numeracy	0.06	0.055	1	457	0.1741	0.676711	
ILT	0.01	0.007	1	457	0.0229	0.879902	
GCB	16.45	16.45	1	457	51.8097	2.53E-12	***
AOT	0	0.001	1	457	0.0041	0.94918	
Social Media Use	0.06	0.061	1	457	0.1911	0.662234	
Political Orientation	21.81	21.809	1	457	68.6896	1.29E-15	***
News Knowledge	0.39	0.386	1	457	1.2161	0.270703	
Familiarity	0.02	0.021	1	457	0.0659	0.797483	
Condition:Stimulus Class	8.02	0.802	10	2330	2.5253	0.005049	**

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Study 2 ANOVA Table for Discrimination

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)	
CRT	12.06	12.06	1	2802.8	34.2983	5.28E-09	***
Numeracy	15.189	15.189	1	2802.3	43.1977	5.88E-11	***
ILT	1.631	1.631	1	2802.8	4.6371	0.03137	*
GCB	40.839	40.839	1	2803.5	116.144	< 2.2e-16	***
AOT	7.811	7.811	1	2802.1	22.2129	2.56E-06	***
Social Media Use	0.994	0.994	1	2802	2.8274	0.09278	.
Political Orientation	59.45	59.45	1	2803.3	169.0721	< 2.2e-16	***
News Knowledge	0.027	0.027	1	2803.7	0.0763	0.78233	
Familiarity	0.306	0.306	1	2802.9	0.8704	0.35092	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Study 2 ANOVA Table for Response Bias

Type III Analysis of Variance Table with Satterthwaite's method

	Sum Sq	Mean Sq	NumDF	DenDF	F value	Pr(>F)	
CRT	0.028	0.028	1	2803.9	0.1443	0.70407	
Numeracy	0.285	0.285	1	2803.1	1.4553	0.22779	
ILT	0.002	0.002	1	2803.9	0.0108	0.91727	
GCB	37.691	37.691	1	2802.8	192.1945	< 2e-16	***
AOT	0.197	0.197	1	2802.3	1.0059	0.31597	
Social Media Use	0.242	0.242	1	2802.1	1.2348	0.26658	
Political Orientation	46.943	46.943	1	2803.7	239.3738	< 2e-16	***
News Knowledge	1.028	1.028	1	2801.6	5.2411	0.02213	*
Familiarity	0.292	0.292	1	2804	1.4884	0.22256	

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1