A Signal Detection Theory Approach for Predicting Immunity to Pandemic Vaccine Fake News

Jeremy Barajas
University of Pennsylvania
barajas@sas.upenn.edu

Richard John
University of Southern California
richardj@usc.edu

Abstract

Managing a contagious disease pandemic, such as COVID-19, requires that the public understand and cooperate with behavioral guidelines to reduce viral transmission. This research uses Signal Detection Theory (SDT) to explore how U.S. adults distinguish between true and false information related to COVID-19 vaccines. A total of 372 U.S. adults categorized 17 true and 17 false COVID-19 vaccine headlines. Item Response Theory analyses suggest that the ability to identify true information about the pandemic is a construct that can be reliably measured with our novel methodology. Signal Detection Theory analyses indicate high accuracy (AUC = 0.861*), with no bias favoring either true or false responses. Overall, U.S. adults correctly classified 7 of 10 true and 8 of 10 false headlines. Multiple regression analyses on individual performance metrics reveal substantially lower accuracy among conservatives and those with lower scores on a measure of Actively Open-minded Thinking. Implications and limitations of these findings within the pandemic news context are discussed.*

Keywords: COVID-19, News Headlines, Bias, ROC Analysis, Truth Signal

1.1 Fake news and behavior

Greene & Murphy (2021) investigated how COVID-19 misinformation impacted individuals' behavior. The researchers created fake narratives about the pandemic and presented them to participants and surveyed their willingness to engage in their associated protective behaviors. Results indicated that participants exposed to the narrative about the contact tracing app reported a 5% reduction in motivation to download the app and engage in the behavior.

Another study found U.S. political conservatism was strongly linked to weaker protective behaviors compared to liberal-leaning individuals (Pennycook et al., 2022). Furthermore, conservative-leaning individuals tended to have lower risk perceptions of COVID-19, and vaccine hesitancy was more pronounced among this group. These results are significant because they establish that misinformation can nudge behavior. Because management of a deadly pandemic hinges on providing the public with accurate and up-to-date information, it is helpful to understand the cognitive mechanisms propelling the belief in COVID-19 misinformation.

Interestingly, the public's struggle with medical misinformation is not new. One study constructed a mental model of parents' understanding of child vaccinations (Downs et al., 2008). Results indicated that many parents lacked proper knowledge of vaccines, making them a target for misinformation. Some of these parents used the internet to get more information about vaccines. The use of simple search terms often led these parents to anti-vaccination sites. Parents who used simple search terms tended to be individuals who needed the information the most, making them the most vulnerable to being influenced by misinformation. The results of these studies demonstrate the need for studies investigating the impact of COVID-19 misinformation on the general public.

A hallmark of effective crisis management is the public's belief and compliance with emergency orders. Investigating the public's relationship with misinformation and information about coronavirus vaccines is critical to understanding how effective crisis management efforts were across the United States.

1.2 The need for a new methodology

Most prior misinformation studies have relied exclusively on the percentage of items participants correctly identified as the sole indicator of truth detection ability. Unfortunately, the percent correct approach severely limits the interpretation of results. Specifically, it does not provide information on thresholds, assumes true and false information is assessed in the same way (which prior research has demonstrated to be untrue) and fails to account for item difficulty. Consequently, we adopted an SDT framework to assess individuals' truth detection ability independent of item difficulty.

A previous study conducted a re-analysis of prior misinformation studies and is particularly relevant to the current research (Batailler et al., 2022). The authors argued there are two essential domains previous misinformation studies (the percent correct approach) conflate: 1) the ability to detect true and

URI: https://hdl.handle.net/10125/102857
978-0-9981331-6-4

false information and 2) response biases to judge news as accurate versus fake, regardless of the veracity of the news. Furthermore, these two constructs are influenced by differing sets of cognitive processes. Therefore, any studies using the percent correct approach only capture a piece of the underlying story regarding why people fall for fake news, leading these studies to unfounded conclusions. The authors argued for integrating SDT analyses into misinformation studies to detangle individual accuracy from response biases (Batailler et al., 2022).

For example, consider a participant tasked with categorizing true and false news headlines (typical for misinformation studies). The participant was exceptional at classifying the true news items and performed poorly when classifying the false news items. Researchers using the percent correct approach would conclude the participant is great at detecting true information and not so great at detecting false information. However, is this a correct conclusion? One could argue the participant possessed a response bias giving them the tendency to accept newly-presented information as true. In other words, it is impossible to determine whether the participant is highly accurate when classifying true information or they are biased towards accepting new information as true. The percent correct approach provides no way to detangle these two constructs (Batailler et al., 2022).

Addressing the concerns outlined in Batailler et al.'s (2022) study, we adopted an SDT framework that provides greater insight into individual truth detection ability. A previous study in our lab provided an account of SDT under the context of social media posts (Byrd & John, 2022). They posited that any post is associated with a latent truth signal. Specifically, to classify a post as true, the participant must perceive the strength of this latent truth signal as above a threshold for classifying posts as true. This criterion is referred to as the individual's threshold. If this latent truth signal is perceived as lower than the participants' threshold, they would be inclined to classify the social media post as false. While each item has a fixed latent truth signal, individuals may perceive different strengths of the signal. Furthermore, truth thresholds may differ across individuals. Byrd and John (2022) used SDT to estimate sensitivity and specificity, where sensitivity was defined as the probability of detecting true information when given true information, and specificity was defined as the probability of detecting false information when given false information.

Byrd and John's (2022) publication emphasizes the point that, following an extreme event, the communication of crisis information is vital. The authors described a dynamic process where the public receives updates from police departments, public health officials, and other crisis entities, and the public can provide information to authorities providing insight into urgent needs. Social media is a powerful tool for facilitating this process. However, the authors note there is an especially low barrier to entering social media. As a result, the potential arises for the rapid spread of false information, resulting in botched management of the crisis resulting from public belief in false information shared on social media and rejection of true information.

Byrd and John (2022) provided an account of SDT analysis whereby all respondents were associated with a unique threshold for classifying true and false information. Individual "hit" rates and "false-alarm" rates can be derived contingent on an individual's threshold. They defined a hit as correctly classifying true information as true and a false alarm as classifying information as true when false. Participants' responses are rooted in both the perceived strength of the latent truth signal from the item and the participants' threshold for classifying items as true or false (Byrd & John, 2022). There are numerous possible combinations of hit and false alarm rates depending on an individual's threshold. Using SDT analysis, a Receiver Operating Characteristic (ROC) can be generated to plot all possible combinations. The area under the curve (AUC) ranges from 0.5 (chance performance) to 1.0 (perfect performance) and represents overall classification performance.

Relying exclusively on participants' percent correct classification does not provide enough information to make evidence-based claims about how well participants can discern true and false information. The integration of SDT into misinformation studies allows researchers to gain a more comprehensive insight into the abilities of their participants to distinguish between true and false information. The current study is novel because it adopted an SDT framework to develop two sets of items, one containing true information and the other containing false information, measuring individuals' ability to classify true and false headlines about COVID-19 vaccines.

Pennycook et al.'s (2022) practical guide for conducting misinformation studies does not explicitly advocate for the integration of SDT analyses, despite a large body of their work covering the measurement and differentiation of motivated reasoning, lazy thinking, and bias. Furthermore, Batailler et al.'s (2022) re-analysis of Pennycook and Rand's (2019) data showed that the belief in fake news is driven by partisan bias and lazy thinking, a much more nuanced conclusion only possible through SDT analyses.

1.3 The current study

The present study is particularly timely as it used news items about the ongoing pandemic and investigated the public's beliefs about the safety and efficacy of coronavirus vaccines. Furthermore, this study attempted to capture real-life social media feeds by including media stories that were trending in real-time. Understanding how certain information resonates with the general public is critical to communication during times of emergency, such as a global health crisis.

The present study develops a novel methodology that measures individuals' ability to detect true and false information about the COVID-19 pandemic. The current research addresses four questions: 1) How difficult is it to discern a typical headline about COVID-19 vaccines, 2) How well can individuals distinguish between true and false information 3) What response biases does the general public possess when encountering novel information, and 4) Are there any cognitive traits within individuals that are predictive of their likelihood of falling for fake news?

We identified two sets of news items using coronavirus vaccine-related headlines: one with true information and the other with false information. We used Item Response Theory (IRT) analysis to estimate item difficulty and discriminability. The 2-parameter model IRT analysis provided estimates of each item's difficulty and ability to discriminate among individuals based on their underlying latent ability to classify true and false information. The IRT analysis provided a basis for removing news items exhibiting lower discriminability.

We utilized a Signal Detection Theory (SDT) framework to characterize both group and individuals performance in classifying true and false information, independent of any bias to say true or false. We also measured several individual difference variables (political orientation, open-mindedness, education, etc.) to predict accuracy, sensitivity, specificity, and bias.

1.4 Individual difference variables

Of interest to this project was how specific psychological characteristics, such as political orientation, are related to individual SDT classification performance metrics. One study investigated this phenomenon, indicating that participants who preferred Donald Trump over Hillary Clinton in the 2016 U.S. presidential election performed worse on the news media task than those who preferred Hillary Clinton (Pennycook & Rand, 2019). Furthermore, the results of Lutzke et al.'s (2019) study demonstrated that politically conservative individuals were more likely to fall for fake news about climate change and mistrust factual information than liberal-leaning participants.

More recently, Calvillo et al. (2020) reported that political conservatism was negatively correlated with the ability to discern true and false information about the pandemic. Their results also indicated that political conservatism was negatively associated with the perceived threat of the Coronavirus.

Another variable of interest is Actively Open-minded Thinking (AOT). AOT is described as a set of standards for evaluating the thinking of others (Baron, 2018a). Baron (2018b) described AOT in a two-steps framework: one's search for knowledge and their response to the interferences arising from the search for knowledge. Within the AOT framework, "interferences" refers to any instances of my-side biases and premature conclusions activated during the search for knowledge (Baron, 2018b).

Applying this construct to the COVID vaccine domain, an individual searching for information about the vaccine would be engaging in a search for knowledge. If this person had a prior belief of the vaccine being dangerous, then the interferences would be the urge to end their search early once they came across an article claiming the vaccine is dangerous. This individual would also be predisposed to discard or ignore information suggesting the vaccine is safe. To engage in Actively Open-minded thinking means conducting an exhaustive search for information and ignoring interferences before the search has concluded (Baron, 2018b).

It is important to note that AOT does not refer to a state of mind and does not require any domain-specific knowledge (Baron, 2018b). Rather, it describes one's disposition to engage in exhaustive searches for knowledge and avoid interferences. Individuals high in AOT experience reduced confidence and place a higher value on additional information, motivating them to seek further information and display an openness to opposing arguments (2018b).

AOT is not to be conflated with critical thinking, which often encourages a general skepticism towards the information itself. AOT encourages the individual to evaluate the thinking process of other sources of information, whether they be other individuals, articles, or internet posts. Baron (2019a) suggests that by engaging in AOT, individuals will have a better sense of the veracity of information without additional domain-specific knowledge.

One related study investigated how AOT interacts with individuals' behavior during the COVID-19 pandemic (Cohen et al., 2021). The results of the study indicated that those higher in AOT indicated greater trust in public health experts, perceptions of higher risks associated with COVID-19, and greater

compliance with protective behaviors (Cohen et al., 2021). The current study investigated the relationship between AOT and U.S. adults' ability to detect true and false information about coronavirus vaccines.

Another critical dimension of truth detection ability is individuals' skepticism when presented with information. Byrd and John (2021) investigated how skepticism is related to the ability to discern true and false information. Participants were presented with true and false social media posts after a natural disaster or a soft-target terrorist attack. Skepticism was measured using the 30-item Professional Skepticism Scale Hurtt (2010). The results indicated that skepticism was a significant predictor of detecting true information for true social media posts about terrorist attacks. We further explored the relationship between skepticism and classification performance in the current study.

Teunisse et al. (2020) reported a series of five studies aiming to develop a reliable measure of gullibility. Their results indicated that those who reported clicking on the link (dubbed a scam victim) tended to score higher on the gullibility scale than those who did not fall for the scam. Furthermore, participants who scored higher on the gullibility scale tended to perceive the scam emails as more persuasive than those who scored lower. The present study utilized their gullibility scale as a possible predictor of truth detection ability.

2. Hypotheses

This study addresses the following six hypotheses:

1. An underlying (latent) ability to identify true and false information related to COVID-19 vaccines can be reliably measured.
2. Due to the intense interest in and scrutiny of pandemic news over the last two years, U.S. adults will demonstrate higher accuracy (AUC > 0.70) for categorizing true and false vaccine stories than typically reported in previous studies.
3. Following previous research reporting a relationship between political identity and vaccine beliefs, we expect that self-reported political orientation will predict overall accuracy (AUC), with conservative-leaning individuals demonstrating poorer performance than liberal-leaning individuals.
4. Following previous research, we expect that Actively Open-Minded Thinking (AOT) will predict overall performance, with individuals higher in AOT performing better than individuals lower in AOT.
5. Following previous research, we expect skepticism to predict accuracy, with individuals higher in skepticism performing better than those lower in skepticism.
6. Following previous research, we expect gullibility to predict accuracy, with individuals higher in gullibility performing worse than those lower in gullibility

3. Methodology

An overview of the experimental methodology and data modeling is presented in Figure 1 below.

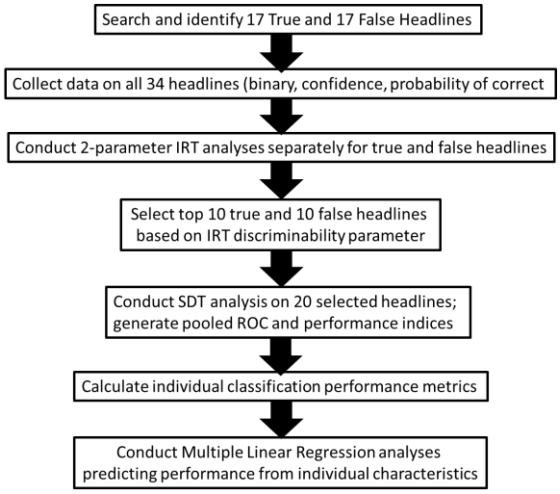

Figure 1. Overview of 7-step methodology.

3.1 Participants

A sample of U.S. adults representing the general population for gender, age, and race (N=375; *M* age = 45 years, *SD* = 16 years) was recruited from Prolific.com. Participants were paid $2 to complete the survey. Three participants were excluded for completing the survey too quickly (less than 9 minutes) or failing attention check questions (0.8% dropped).

A majority of participants (50.5%) self-identified as female, while 48.1% of participants identified as male, which was collected via the slider function on Qualtrics. Participants' political ideology was collected using a 7-point scale ranging from 1= extremely liberal to 7= extremely conservative. Furthermore, a majority of participants reported having a bachelor's degree (37.1%), followed by participants who attended some college (27.7%), then participants who reported only having a high school degree (16.9%), and 16.4% of participants reported having a graduate or professional degree.

In terms of political ideology, 143 (38.2%) were extremely or very liberal, 65 (17.5%) were slightly liberal, 85 (22.8%) were moderate, 34 (9.1%) were slightly conservative, and 46 (12.3%) were extremely/very conservative. In terms of racial/ethnic

makeup, 260 (69.9%) were Non-Hispanic White or Euro-American, 52 (14.0%) were Black, Afro-Caribbean, or African American, and 15 (4.0%) were Latino or Hispanic.

3.2 Procedure

Respondents were presented with 34 media headlines, randomly ordered, from various media platforms and news outlets via the Qualtrics survey platform. Seventeen headlines contained false information, as determined by Snopes.com and Politifact.com. The remaining seventeen headlines included true information obtained from reputable outlets such as The Associated Press, NPR, and Reuters. Each media headline depicted a fact or event related to coronavirus vaccines. At the beginning of the survey, respondents were told that each headline contained either true or false information, but they were not told the proportion of true and false headlines. Respondents were asked to make a dichotomous judgment about the content of each headline (true or false), rate their confidence in their judgment (a 5-point scale ranging from not confident at all to very confident), and finally, rate the likelihood that their judgment is correct using a slider scale from 50 to 100 (where 50 indicates a guess and 100 indicates certainty).

After completing all 34 headline items, participants completed the AOT, skepticism, and gullibility questionnaires. These measures presented the participant with a series of statements, and the participant indicated their agreement with each statement by using a Likert scale. Finally, participants completed a demographic questionnaire and were asked to provide feedback about anything they found confusing throughout the survey. The median time to completion was 12 minutes.

3.3 Media Story Section

The headlines were sourced from the following websites: Six from Snopes, eight from the Associated Press, 14 from Reuters, four from Axios, one from Politifact, and one from USA Today. Examples of four true and four false headlines are presented in Table 1. Thirty-four headlines depicting a fact or an event about coronavirus vaccines were selected. The 17 false headlines originated from websites such as Snopes.com or Politifact.org, as they have sections of their sites dedicated to fact-checking trending claims about the Coronavirus. The 17 true headlines were obtained from reputable publications such as the Associated Press, Snopes, Reuters, NPR, and Pew. Snopes and Politifact can label claims anywhere from completely false, mostly false, or a mixture of true and false. The current study included false headlines flagged as completely false only. Following IRT analysis, seven true and seven false items were removed for low discriminability, resulting in ten true and ten false news items.

Table 1. Example true (top) and false (bottom) headlines.

	Date:	Example True Headlines:
1	21-Aug-21	Though young and healthy, unvaccinated father dies of COVID: He was completely healthy, active, not a smoker
2	29-Jun-21	Nearly all COVID deaths in US are now among unvaccinated.
3	14-Jun-21	Novavax: Large study finds COVID-19 shot about 90% effective.
4	24-Jun-21	Pfizer says COVID vaccine is highly effective against Delta variant.
	Date:	Example False Headlines:
1	17-Jun-21	Three British Airways Pilots Have Died of the COVID "Vaccine" in the Past 7 Days.
2	22-Apr-21	REPORT: Pfizer Vaccine Confirmed to Cause Neurodegenerative Diseases - Study
3	21-Jul-21	mRNA Shots Caused Miscarriages In 82% OF Women – One U.K. study shared that 104 of 127 pregnant women faced miscarriages after the Pfizer and Moderna mRNA shots.
4	6-Aug-21	COVID-19 mRNA vaccines contain graphene oxide.

3.4 Individual difference measures

The 7-item AOT scale described in Haran (2013) was used. Skepticism was measured using a shortened version of the 30-Item *Professional Skepticism Scale* found in Hurtt (2010). Participants read each statement and indicated their agreement with it using a 6-point Likert scale, where a higher score indicates a higher degree of skepticism. Gullibility was measured using the 12-item *Gullibility scale* Teunisse et al. (2020). Participants read each statement and indicated their agreement with the statement using a 7-point Likert scale, where higher scores indicate a higher degree of gullibility.

3.5 Analysis overview

Three phases of data analysis were conducted. In the first phase, IRT analysis was conducted to yield estimates of discriminability and difficulty for each news item. In the second phase, Signal Detection Theory (SDT) was applied to the pooled set of headline items to construct a Receiver Operating Characteristic (ROC) curve that provides estimates of classification performance metrics (AUC, d', A', c, sensitivity, and specificity) at both the group and the

individual level, In the third phase, Multiple Linear Regression (MLR) was conducted to predict individual respondents' accuracy, bias, sensitivity, and specificity from the various demographic variables and the psychometric measures of cognitive functioning.

4. IRT Analysis of Headline Items

The current study utilized the two-parameter IRT model, where difficulty and discriminability are estimated for each headline (item). Item difficulty is defined as the latent ability level corresponding to a 50:50 chance of classifying the headline correctly. Discriminability is defined as the degree to which the likelihood of correct classification is different for those just above and just below the ability level corresponding to item difficulty (An & Yung, 2014; Embretson, & Reise, 2000).

Generally, items with high discriminability are preferred, as items low in discriminability are not useful in measuring ability to classify items as true or false. Operationally, the Item Characteristic Curve (ICC) graphically defines both difficulty and discriminability (An & Yung, 2014; Embretson, & Reise, 2000). The difficulty for a particular item is defined as the ability level corresponding to the intersection of the ICC and the p=0.50 probability of correct classification. Likewise, discriminability is defined by the steepness of the slope of the ICC at the ability level corresponding to item difficulty (An & Yung, 2014; Embretson, & Reise, 2000).

Separate 2-parameter IRT models were estimated for the 17 true and 17 false items using the "ltm" package in R (Rizopoulos, 2006). Each model includes both a difficulty and a discriminability parameter for each item. Estimated parameter values and ICC curves plotting the probability of identifying the headline as true (or false) vs. a latent ability construct were inspected. Seven true items and seven false items were dropped based on low discriminability. Following removal of the 14 items, IRT parameters were again estimated with the remaining 20 items.

As seen in Figure 2, each ICC represents a different headline (item). The vertical axis indicates the probability of correctly classifying each headline, contingent upon the ability level of each participant. The steepness of the curve near the intersection with the p=0.50 horizontal line represents the item's discriminability, where steeper indicates higher discriminability.

As evidenced by the curves in Figure 2, the most discriminating items appeared to be low to medium difficulty for the nationally-representative sample. This is evidenced by a 0.5 chance of correctly identifying the item being just below 0 on the ability axis for both true and false items. Note that no ICC plots are flat, indicating that the likelihood of correctly categorizing an item increases as individual ability increases.

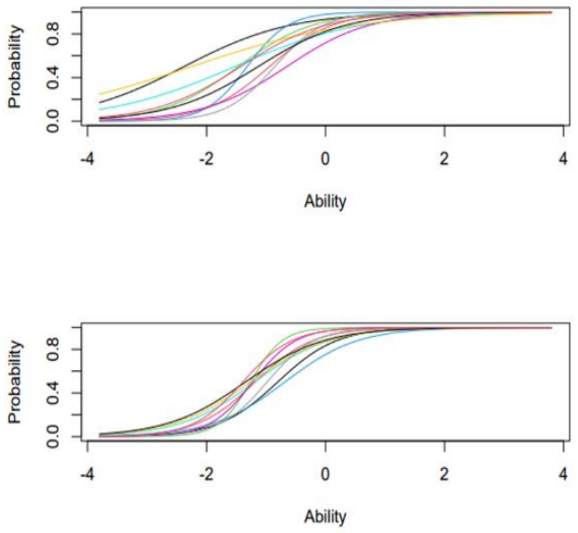

Figure 2. Item Characteristic Curve (ICC) Graphs for 10 True (top) and 10 False (bottom) Vaccine Headlines.

Test Information Function (TIF) curves indicate the overall ability of the entire set of items to differentiate individuals based on their ability to detect true and false information. As evidenced by Figure 3 below, both TIF curves are slightly to the left of 0 for latent ability, indicating the news items are of lower to medium difficulty. The TIFs still demonstrate that the items can discriminate to a greater or lesser extent across a range of ability levels.

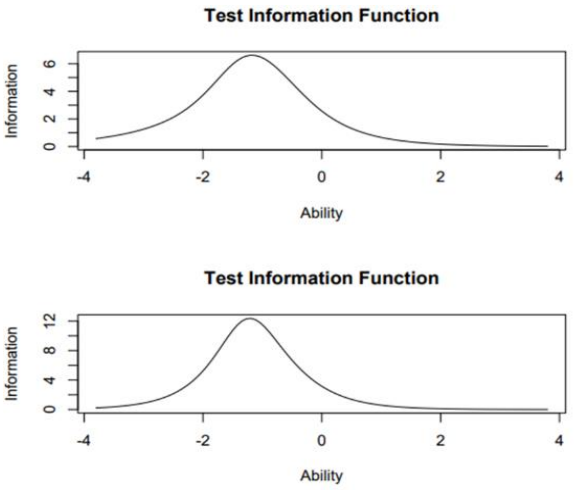

Figure 3. Test Information Function (TIF) Curves for 10 True (top) and 10 False (bottom) Vaccine Items.

5. Pooled SDT analysis

Responses were pooled across all 372 respondents and 20 responses for each respondent to estimate a pooled ROC curve. Following previous applications of SDT to analyze accuracy in categorizing true and false news headlines (Byrd & John, 2022). We unfolded the dichotomous responses and confidence ratings to a continuous scale ranging from 1 (low truth signal) to 10 (high truth signal) using the following formula:

Score = 5 + Confidence, for headlines judged true
Score = 6 – Confidence, for headlines judged false

We also unfolded the likelihood of correct (L.C.) true or false judgments, ranging from 50 (guessing) to 100 (certainly correct), to create a continuous probability of true variable ranging from 0 (certainly false) to 100 (certainly true) using the following formula:

Likelihood = L.C., for headlines, judged true
Likelihood = 100 – L.C., for headlines judged false

Both confidence scores and likelihood judgments were used to create separate estimates of the ROC curve. The two dashed curves plot sensitivity vs. false alarm rate (= 1 – specificity) for the pooled sample of responses. The ROC curves in Figure 4 were constructed independently using confidence ratings and probability judgments of correct classification and are virtually indistinguishable. The diagonal reference line represents chance performance. The estimated AUC for the ROC curve using confidence ratings was 0.861, with 95% CI = (0.855, 0.868); the AUC for the ROC curve using probability of correct judgements was 0.863, with a 95% CI = (0.857, 0.869). The AUC can be interpreted as the probability that a randomly selected true item headline will have a greater truth score than a randomly selected false item headline for a randomly selected respondent.

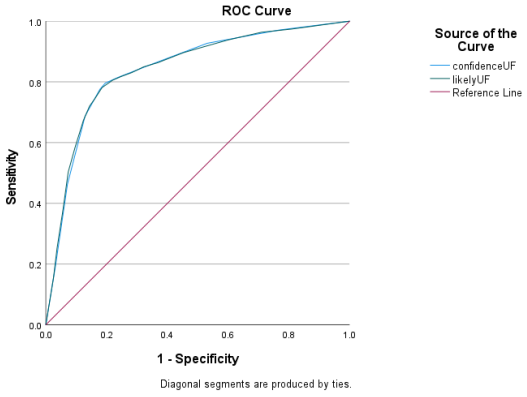

Figure 4. Pooled Receiver Operating Characteristic (ROC) curves (dashed) plotting sensitivity (vertical axis) vs. 1-specificity (false positive rate).

6. Individual differences in accuracy

In order to investigate hypotheses related to individual differences in accuracy and bias, SDT parameters were calculated separately for each respondent. We used the binary responses for individual analyses only, which corresponds to the 2AFC (2 alternatives, forced-choice) method (Stanislaw & Todorov, 1999). This approach necessarily assumes that the underlying distributions for the latent truth signal for true and false headlines are Normal with equal variance. Both d' and A' are overall measures of individual accuracy; d' represents a normalized measure of the mean difference between true and false headlines, and A' is an estimate of the area under the ROC curve, conditional on Normal, equal variance latent distributions.

Sensitivity and specificity are the proportion of correctly identified true and false headlines, respectively. The parameter c is a measure of bias and is calculated as the difference between the optimal threshold and each individual's observed empirical threshold (beta), again conditional on the assumption that the latent construct for both true and false headlines have Normal distributions with equal variance. Finally, the proportion of true responses is also considered a measure of bias. Instructions to maximize the number of correct responses and that half the headlines are true and half false indicates that deviations from a 50% response true response rate imply bias favoring either true or false responses.

Table 2. Means, standard deviations, and 95% confidence intervals for individual SDT parameters, N=372.

SDT Parameter:	Mean	SD	95% CI Lower	95% CI Upper
d'	1.895	1.342	1.76	2.03
A'	0.845	0.225	0.82	0.87
Sensitivity	0.796	0.216	0.77	0.82
Specificity	0.804	0.254	0.78	0.83
c (bias)	0.020	0.375	-0.02	0.06
True Response Rate	0.496	0.114	0.48	0.51

Table 2 presents means, standard deviations, and 95% confidence intervals for all six SDT parameters

calculated for each individual. Mean d' is different from 0.0 (complete overlap of the latent true and false distributions), p<0.001. Likewise, A' is different from 0.50 (chance accuracy), p<0.001. Note that the value of the mean A', based on the Normal distribution, equal variance assumption, is quite close to the value of AUC calculated from the pooled ROC curve. The accuracy measures (d', A', sensitivity, and specificity) show moderate variation across individual respondents. It is noteworthy that the mean of c (bias) is not significantly different from 0.0; likewise, the base rate of true responses is not significantly different from 50%.

We examined the individual differences in overall accuracy by conducting a multiple linear regression predicting d' and A' from political orientation (P.O.), age, education, AOT, skepticism, and gullibility. Table 3 reports the raw regression parameters and standard errors for each predictor of d' and A'. The predictors accounted for variance in d' and A' significantly greater than zero; Multiple R=0.575 for d', (F(7, 364)=25.647, p<0.001) and R=0.540 for A', (F(5,364)=21.439, p<0.001). As indicated in Table 3, political orientation and AOT significantly predicted both d' and A'. Political orientation was significantly predictive of both d' and A'. None of the other variables tested were significantly predictive of accuracy.

Table 3. Raw regression parameters and standard errors for predicting d' and A', N=372.

	d'		A'	
Predictor:	B	S.E.	B	SE
Constant	1.559	0.831	**0.806**	0.143
PO (Political Orientation)	**-0.332**	0.040	**-0.051**	0.007
Age	0.010	0.004	0.001	0.001
Education	0.013	0.041	0.005	0.007
AOT	**0.380**	**0.077**	**0.061**	0.013
Skepticism	-0.240	0.149	-0.043	0.026
Gullibility	-0.075	0.079	-0.004	0.014

Notes: **Bold,** p < 0.001

The regression parameters reported in Table 3 suggest that each unit increase from liberal to conservative political orientation decreased d' and A' by nearly -0.332 and -0.051, respectively. In addition, each unit increase in AOT resulted in increases in d' and A' by 0.380 and 0.061, respectively.

Individual differences in accuracy were further investigated using multiple linear regression predicting sensitivity and specificity. The predictors accounted for variance in sensitivity and specificity significantly greater than zero; Multiple R=0.476 for sensitivity, (F(7, 364)=15.222, p<0.001) and R=0.534 for specificity (F(7, 364)=20.709, p<0.001). As indicated in Table 4, political orientation and AOT significantly predicted both sensitivity and specificity.

Table 4. Raw regression parameters and standard errors for predicting sensitivity and specificity, N=372.

	Sensitivity		Specificity	
Predictor:	B	S.E.	B	S.E.
Constant	0.**664**	0.144	0.**825**	0.162
PO (Political Orientation)	**-0.045**	0.007	**-0.055**	0.008
Age	0.002	0.001	0.001	0.001
Education	-0.003	0.007	0.008	0.008
AOT	0.**049**	0.013	0.**067**	0.015
Skepticism	-0.026	0.026	-0.048	0.029
Gullibility	0.008	0.014	-0.024	0.016

Notes: **Bold**, p < 0.001

Regression parameters in Table 4 estimate that each unit increase from liberal to conservative political orientation decreased sensitivity and specificity by -0.045 and -0.055, respectively. Furthermore, each unit increase in AOT reduced sensitivity and specificity by 0.049 and 0.067, respectively.

Regression analyses on bias (c) and true response rate indicated that neither was related to the individual difference variables.

7. Discussion

The current study was an analysis of misinformation, fake news, and public health communications as perceived by the public. The results of the current study shed light on the efficacy of public health officials' mitigations (risk communications). Understanding how the public responds to new information about coronavirus vaccines addresses the public's pandemic preparedness. The IRT analysis allowed selection of items that form a measure of a latent construct of ability to correctly categorize true and false information about a severe global health crisis. Most items were found to be "easy" or "moderately difficult" in the IRT analysis, which is likely a consequence of the intense scrutiny associated with all pandemic-related information in the media.

In the current study, which used a nationally representative sample of U.S. adults, AUC was 0.861, indicating that the underlying truth signal for true

headlines is greater than for false headlines in 86% of comparisons of randomly selected pairs of true and false headlines. The estimates for sensitivity and specificity were .664 and .825, respectively. While observed accuracy is high, it is important to consider the practical implications of observed errors.

For example, a U.S. adult who reads two weekly headlines would be expected to read 52 true and 52 false headlines over a year, assuming equal proportions of true and false information. The SDT results suggest that 30% of the true headlines and 20% of the false headlines would be classified incorrectly, resulting in rejecting 16 true headlines as false and accepting ten false headlines as true over a year. Even with the relatively high accuracy observed, there is ample opportunity to miss important true information and be misled by false information over a year. Negative health impacts, such as severe illness or death, can be expected to follow from rejection of true information and acceptance of false information.

Substantial variability in accuracy was observed in the sample of U.S. adults. Self-reported political orientation was predictive of overall accuracy. There were remarkably large differences in accuracy between very conservative and very liberal adults.

As political conservatism increases, the ability to correctly identify true and false information about COVID-19 decreases, and U.S. adults' self-reported political orientation was predictive of overall accuracy. As political conservatism increases, the ability to correctly identify true and false information about COVID-19 decreases.

Open-mindedness (AOT) was also significantly predictive of accuracy. Individuals' ability to discern true and false information increases as AOT increases. Furthermore, age, education, skepticism, and gullibility were not significantly predictive of accuracy.

Bias in setting the threshold for responding true vs. false was negligible. Bias is usually a consequence of respondents perceiving one or more of the following: (1) either true or false information is more prevalent (base-rate other than 50:50), or (2) perceiving unequal error penalties for misclassifying a false headline as true or a true headline as false. This finding is consistent with previous research indicating that even explicitly specified base rates and error penalties are under-utilized in setting a threshold for responding true or false (Byrd and John, 2022).

One should not have to be an expert in infectious diseases to remain safe throughout a pandemic. That age and education level were not found to be significant predictors of overall accuracy suggests that public health officials were generally successful at communicating pertinent health information regardless of demographic characteristics. However, considering how political orientation was significantly predictive of accuracy, these studies shed light on a group of individuals who didn't get the message, revealing a lacuna in current risk communication strategies. Managing a deadly disease pandemic requires all individuals to understand and believe in public health communications and guidelines. Actively open-minded thinking presents a new direction for risk communication experts to explore when developing novel ways to ensure information is accepted across all demographics.

8. Limitations

It is important to note several limitations in the present study:

1. These studies use a diverse sample of headlines from the COVID-19 pandemic, which occurred during a highly politicized period. Results may or may not generalize to other health crises, particularly those with fewer political narratives.
2. The classification of headlines as true or false is based on research conducted by the sources cited in Section 3.2. All factual determinations are subject to the acquisition of new evidence, and no fact-check can be 100% guaranteed to be correct as new evidence is discovered. Therefore, it is certainly possible that some true items will not be entirely true in the future, and some false items will not be wholly false.

9. Future research

Future research should explore how the measure of truth discernment ability generalizes to other domains, such as climate change, monkeypox, and other crises. The coronavirus pandemic was accompanied by an unprecedented wave of changing information from a multitude of sources. Therefore, there is a major question of generalizability in less extreme contexts.

Future research should investigate how health communications can be designed with a stronger truth signal to see how it impacts vaccination intentions and engagement in other protective behaviors. Knowing this, public health officials can design their communications to motivate the general public to differentiate between true and false information and act in their own best interest, as well as that of the community.

10. Conclusions

There are six conclusions from the current study:

1. U.S. adults have a measurable, underlying (latent) ability to correctly identify both true and false media information related to the COVID-19 pandemic.
2. U.S. adults demonstrated high accuracy in distinguishing true and false vaccine-related headlines.
3. U.S. adults exhibited little bias toward either true or false responses.
4. Political orientation was highly predictive of all accuracy indices; conservative adults were substantially less accurate than liberal adults.
5. Open-mindedness (AOT) was predictive of all accuracy indices; people who exhibited less open-mindedness were substantially less accurate than those who showed higher degrees of open-mindedness.
6. Age, education level, skepticism, and gullibility were not predictive of overall accuracy in identifying true and false headlines.

11. References

An, X., & Yung, Y. F. (2014). Item response theory: what it is and how you can use the IRT procedure to apply it. *SAS Institute Inc. SAS364-2014*, *10*(4).

Batailler, C., Brannon, S. M., Teas, P. E., & Gawronski, B. (2022). A Signal Detection approach to understanding the identification of fake news. *Perspectives on Psychological Science*, *17*(1), 78–98. https://doi.org/10.1177/1745691620986135.

Baron, J., 2019a. Actively open-minded thinking in politics. *Cognition*, *188*, pp.8-18.

Baron, J. (2018b). Individual mental abiities vs. the world's problems. *Journal of Intelligence*, *6*(2), 23. https://doi.org/10.3390/jintelligence6020023

Byrd, K., & John, R. S. (2022). Lies, damned lies, and social media following extreme events. *Risk Analysis*, *42*(8), 1704-1727. https://doi.org/10.1111/risa.13719.

Byrd, K., & John, R. (2021, January). Tell me the truth: Separating fact from fiction in social media following extreme events. In *Proceedings of the 54th Hawaii International Conference on System Sciences* (p. 2718). http://hdl.handle.net/10125/70946

Calvillo, D., Ross, B., Garcia, R., Smelter, T., & Rutchick, A. (2020). Political ideology predicts perceptions of the threat of covid-19 (and susceptibility to fake news about it). *Social Psychological & Personality Science*, *11*(8), 1119–1128. https://doi.org/10.1177/1948550620940539

Cohen, A. S., Lutzke, L., Otten, C. D., & Árvai, J. (2021). I think, therefore I act: The influence of xritical reasoning ability on trust and behavior during the COVID-19 Pandemic. *Risk Analysis*. https://doi.org/10.1111/risa.13833

Downs, J. S., de Bruin, W. B., & Fischhoff, B. (2008). Parents' vaccination comprehension and decisions. *Vaccine*, *26*(12), 1595–1607. https://doi.org/10.1016/j.vaccine.2008.01.011

Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for psychologists.* Lawrence Erlbaum Associates: Hillsdale, NJ.

Greene, C. M., & Murphy, G. (2021). Quantifying the effects of fake news on behavior: Evidence from a study of COVID-19 misinformation. *Journal of Experimental Psychology. Applied*, *27*(4), 773–784. https://doi.org/10.1037/xap0000371

Haran, U., Ritov, I., & Mellers, B. A. (2013). The role of actively open-minded thinking in information acquisition, accuracy, and calibration. *Judgment and Decision Making*, *8*(3), 188–201. https://journal.sjdm.org/13/13124a/jdm13124a.pdf

Hurtt, R. (2010). Development of a scale to measure professional skepticism. *Auditing: A Journal of Practice and Theory*, *29*(1), 149–171. https://doi.org/10.2308/aud.2010.29.1.149

Lutzke, L., Drummond, C., Slovic, P., & Árvai, J. (2019). Priming critical thinking: simple interventions limit the influence of fake news about climate change on Facebook. *Global Environmental Change*, *58*, 101964. https://doi.org/10.1016/j.gloenvcha.2019.101964.

Pennycook, G., Binnendyk, J., Newton, C., & Rand, D. G. (2021). A practical guide to doing behavioral research on fake news and misinformation. *Collabra: Psychology*, *7*(1), 25293. https://doi.org/10.1525/collabra.25293

Pennycook, G., McPhetres, J., Bago, B., & Rand, D. G. (2022). Beliefs About COVID-19 in Canada, the United Kingdom, and the United States: A Novel Test of Political Polarization and Motivated Reasoning. *Personality & Social Psychology Bulletin*, *48*(5), 750–765. https://doi.org/10.1177/01461672211023652

Shearer, E. (2021, January 12). *More than eight-in-ten Americans get news from Digital Devices*. Pew Research Center. Retrieved June 2, 2022, from https://www.pewresearch.org/fact-tank/2021/01/12/more-than-eight-in-ten-americans-get-news-from-digital-devices/

Pennycook, G., & Rand, D. (2019). Lazy, not biased: susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, *188*, 39–50. https://doi.org/10.1016/j.cognition.2018.06.011

Rizopoulos D (2006). ltm: An R package for latent variable modeling and item response theory analyses. *Journal of Statistical Software* 17(5):1–25.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments & Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Teunisse, A., Case, T., Fitness, J., & Sweller, N. (2020). I should have known better: development of a self-report measure of gullibility. *Personality & Social Psychology Bulletin*, *46*(3), 408–423. https://doi.org/10.1177/0146167219858641