

Impact of Including a Photo with a News Headline on Veracity Judgements for Fake News

By Maxine Ewing

Dietrich Senior Honors Thesis with Carnegie Mellon University

Professor Gretchen Chapman

2021

Acknowledgements

I would like to thank the insurmountable support that I have received from friends and mentors alike. This especially goes to my thesis advisor, Gretchen Chapman, for her unwavering support, for teaching me so much, and for supporting my growth through this process.

I would also like to thank Aaron Satyanarayana for always answering my questions about R, and for responding to my pleas to read over my writing with humor and grace.

I would like to thank my parents, as always, for listening to me try to explain what this is all about, and for giving me the equipment in life to be prepared for such an undertaking.

Lastly, I would like to thank the many people in my life that have supported me and most importantly, encouraged me, to pursue something new and exciting, and for staying by my side all the way.

ABSTRACT

Fake News comes in many, pervasive forms, and has dominated the conversation around important world events, including the 2016 and 2020 United States presidential election. In this study, we explore how photos as a medium may be impacting how people are able to distinguish real from fake news. Previous research has explored how elements such as news sources, traits and characterizations of people, and system 1 and 2 processing affect how people discern fake from real news. However, the question of how different types of media may be impacting how people read news has yet to be explored. In the current study participants judged the veracity of 30 news article headlines, all of which actually circulated in the US around 2016 to more current times. Inclusion of a photo did not significantly change how well participants discriminated fake from real news, suggesting that people do not use photos as a cue for the believability of a story. Further research is necessary to better understand how photos and other forms of media may impact the believability of news.

1 INTRODUCTION

Fake news is one of the most relevant topics in recent history and a major issue society faces today, especially in regard to elections and conspiracies. The conversation about fake news was at the forefront of, for example, the United States Capitol Riots on January 6, 2021, in which thousands of people protested outside the Capitol, and many rushed inside (Leatherby et al 2021). This is in no way the only example of fake news' firm grip on the polarization of the United States, but it was a deadly example of how big of an issue this really is.

There are many factors that contribute to fake news believability, and there are also many unanswered questions about what makes people fall for fake news. According to a study by Garimella and Eckles (2017) on photo misinformation spreading through WhatsApp, there are four identified categories of image-based fake news: old images out of context, memes (typically with false statistics), manipulated/edited images, and lastly other images that include fake alert and others.

There are two major components in fake news research. The first is one's ability to discriminate between fake and real news, and the other is the believability of fake news. It is possible that different factors will influence these two measures differently. For example, one factor could increase the believability of fake news and decrease discrimination abilities, and vice versa. Existing literature has identified five factors that impact how much people discern fake from real news, and what makes news more believable. The first factor, increased deliberation time over news, makes fake news less believable and does not make real news less believable (Bago et al 2020). This paper also found that increased deliberation time does not result in an increased use of political bias. This 2020 study involved participants answering

whether or not real news headlines were depicting actual events accurately. There was an equal distribution of republican-consistent, democrat-consistent, and fake and true news stories. This was in order to assess how deliberation time and partisanship may interact. Although increasing deliberation time is successful, people don't always spend the time to deliberate when they read news, and there is no currently widespread infrastructure to encourage this. Additionally, fake news outlets aim to maximize believability of their content.

The second factor concerns character traits of individuals. According to Bronstein et al. (2019), believability of fake news is particularly high for dogmatic individuals and religious fundamentalists, likely because people with those traits already hold beliefs that are not evidence-based. Participants were tasked with determining which stories are fake and real out of a distribution of left and right leaning stories. The researchers then measured dogmatism and delusionalism. This article suggests that interventions which increase open-mindedness and analytical thinking could improve discrimination for such individuals.

A third factor is the news source. Clayton et al. (2019) found that for both left leaning and right leaning individuals, the content of the information is more impactful on veracity judgements than the source. In this study, they varied whether or not the same information appeared to come from "Fox News" or "CNN" for the participants. This research could suggest that source overall doesn't play a large role in how people sift through news.

A fourth factor is prior exposure to the news. A 2018 article by Pennycook et al. found that exposure to a headline increases perceived accuracy of the same headline in close temporal proximity or in more distant temporal proximity. Being exposed to news previously makes one more likely to believe it in the future.

The final factor is also explored in a paper by Pennycook et al. (2019). This paper explores how the degree to which individuals ascribe meaning to “pseudo-profound” sentences is correlated with perceived fake news accuracy. That is to say, people who are more likely to believe “bullshit” are also more likely to believe fake news. In one of their studies, they found that people who ascribe more meaning to “pseudo-profound” sentences are less able to discern fake from real news compared to those who are more skeptical. According to the authors, these findings indicate that “open-mindedness” may actually be, in some circumstances, a disadvantage for accurate judgements on news veracity. In conversation with the earlier factor of dogmatism as explored by Bronstein et al. (2019), we are beginning to paint a picture of what types of people may inherently believe fake news more, and thus would benefit most from mitigation strategies.

Existing literature has not explored how different media may influence the discrimination abilities of consumers. Fake news comes in all sorts of forms, from text, to visual, to audio, and people may be influenced in different ways depending on the types of media that are present. Understanding this influence can better inform mitigation strategies for the various types of fake news. In the current study, we explored the question of how the inclusion of an associated image along with a headline might change the veracity judgement of this headline. Because people are constantly being bombarded with news and information, it is likely that they use cues to differentiate fake from real news, beyond their immutable character traits. We wanted to explore how photos may influence news believability. For example, including a photo along with a headline could come across as additional “evidence” to the viewer, thus increasing believability. We predicted that the inclusion of a photo could also be a distraction from the information in the headline, thus coinciding with the findings in Bago et al., which explored deliberation time

(2020). A 2018 paper by Pennycook and Rand found that lazier, less analytical thinking was a significant predictor in partisan fake news perceived accuracy over even partisan beliefs; partisanship is not as much a problem as just a lack of analytical thinking. We hypothesized that the inclusion of a photo would pull veracity judgment scores towards the average, that is, people would rate true things to be more fake, and fake things to be more true, thus signaling a clouding in judgement due to the photo. We were interested in how this effect may differ when the headline matches or does not match a person's political leaning. Consequently, we included the political leaning of participants in our analysis, and we included an even distribution of fake and real headlines that had right, left, or neutral political leanings.

The stimuli we used were borrowed from the 2019 Pennycook et al. paper, with slight modification. In the original paper, participants were presented with thirty headlines, each with an associated image. For our study, half of the participants saw only the headlines, with the photos cut out. The other half saw the complete headline and photo, as originally used in this article. We recruited the participants such that half were democrats, and half were republicans. We used these stimuli because they were actual headlines that had circulated ranging from the 2016 presidential election to more current times, and they had been previously vetted to establish their status as true or fake. Our main dependent variable was a veracity rating question that used the same language as one of the questions in the Pennycook study.

We were only interested in these components, so we did not randomize news source or “genre” of the news (for example, news pertaining to domestic political leaders versus foreign leaders).

2 METHODS

In this study, we used a 2 (real or fake headline) x 2 (photo included or not) x 3 (headline was right-leaning, left-leaning, or neutral) design. The dependent variable was veracity judgements (4 point scale, see appendix for sample question).

Participants: We had a sample size of $n = 200$, and no participants were excluded from the analysis. Participants were recruited from Amazon's Mechanical Turk. In order to recruit equal numbers of democrat and republican participants, we posted two separate HITs on MTurk, one of which was sent to those who identify as Republicans according to MTurk, and one for Democrats, also determined by MTurk. We also included a check question in the survey. 53% participants identified as democrats in this check. We did not exclude participants that did not report an alignment, and conducted analysis based on MTurk's designation as Democrat or Republican.

Headlines: All participants read the same thirty headlines, presented in randomized order. Participants were randomly assigned to a photo or no photo condition. Those in the photo condition read the headlines and were also presented with an associated image. All headlines and images were borrowed from the Pennycook and Rand paper (2019). Ten headlines were right leaning, ten were left leaning, and ten were politically neutral, as determined by the creators of the original stimuli. In each political leaning category, five real headlines contained accurate information, and five contained misinformation or fake news. The headlines came from around 2016. After reading each headline, participants judged its veracity on a 4-point scale.

Survey: In addition to judging veracity, participants were also asked if they were familiar with each headline. On average familiarity participants judged 17% of headlines to be familiar, which indicates that the majority of responses to this question communicated that they were unfamiliar with the headline. We included this familiarity judgement as a control in our analysis because of the implications of Pennycook et al. (2018) concerning repeated exposure.

We also gathered demographic information from the participants: [age, gender, political leaning, education level, race and ethnicity]. **Table 1** in the *Figures* section describes the participants.

87% of participants were white, approximately 50% of participants were over the age of 40, and approximately 93% of participants had a college education or more (**Table 1** in the *Figures* Section).

Data: The veracity judgement was answered by participants on a four point scale (To the best of your knowledge, how accurate is the claim in the above headline? : 1 - Not at all accurate, 2 - Not very accurate, 3 - Somewhat accurate, 4 - Very accurate), as phrased by a Pennycook and Rand study (2019). We averaged the responses across items for each participant to create six lines of data for each participant representing each category of headline. These categories can be found in **Figure 1**. These average ratings comprised a continuous score suitable for regression analysis.

Figure 1: Categories of stimuli

Left True stories
Right True stories
Neutral True stories
Left False stories
Right False stories
Neutral False stories

The study was pre-registered at AsPredicted at <https://aspredicted.org/blind.php?x=yu493i>

3 RESULTS

We ran four pre-registered analyses that describe the results of this study. The planned analyses did not follow the 2x2x3 design because we planned to convert the political leaning of the headline to a variable indicating the match between the headline leaning and the political leaning of the participant. The matching variable is 0 when the story is not in political alignment with the participant (i.e. a democrat reading a right leaning or neutral headline), and is 1 when there is a match (i.e. a democrat reading a left leaning headline).

Analysis 1:

Ground truth refers to the actual veracity of a headline (true or fake). An effect of ground truth on the veracity judgement would indicate that participants can discriminate true from fake headlines - that is, they give higher veracity ratings to true headlines than to fake ones. We examined whether this discrimination was weaker when in the photo condition than in the no photo condition. In all regression analyses, errors were clustered by participant.

We ran a multivariate regression with the following model:

$$Y = B_1 \times \text{Condition} + B_2 \times \text{Ground Truth} + B_3 \times \text{Condition} \times \text{Ground Truth} + \varepsilon$$

Table 2 reports the regression table. The main effect of ground truth indicates that participants gave higher veracity ratings to true headlines than to fake ones. According to the regression model, the interaction between condition (participants who only read the headlines

versus participants who also saw a photo) and the ground truth of the headline (whether or not it was fake or real news) is not significant, showing that discrimination between true vs. fake headlines did not differ between the photo and no photo conditions (see **Figure 2** in *Figures* section). The main effects of “condition” and “ground truth” were both significant at the 5% level of significance. This means that people who saw the photos rated the headlines to be slightly more true, on average, and, as expected, “ground truth” is a good predictor for the veracity judgement, meaning that whether or not the story was fake or real did generally align with how people rated the veracity of the headlines.

Analysis 2:

We repeated the first analysis, now controlling for participant demographic variables:

$$Y = B_1 \times \text{Condition} + B_2 \times \text{Ground Truth} + B_3 \times \text{Female} + B_4 \times \text{Age} + B_5 \times \text{Education} + B_6 \times \text{White} + B_7 \times \text{Condition} \times \text{Ground Truth} + \varepsilon$$

The demographic variables are coded as follows: female: 1= female, other gender = 0; age: 0 is under 40, 1 is 40 and older; education: 0 is below college degree, 1 is college degree or higher; white: white = 1, other race = 0.

None of the demographic variables are significant predictors at the 5% level. As in analysis 1, ground truth affects judgments but there was no interaction between condition and ground truth.

Analysis 3:

We next considered the match variable. We examined a three-way interaction to assess whether the interaction between condition and ground truth occurred when the headline leaning matched the political leaning of the participant. We used the following model for this analysis:

$$Y = B_1 \times \text{Condition} + B_2 \times \text{Ground Truth} + B_3 \times \text{Match} + B_4 \times \text{Condition} \times \text{Match} \\ + B_5 \times \text{Ground Truth} \times \text{Match} + B_6 \times \text{Condition} \times \text{Ground Truth} \\ + B_7 \times \text{Condition} \times \text{Ground Truth} \times \text{Match} + \varepsilon$$

As in the previous analyses, the effect of ground truth indicated that participants gave higher ratings to the true headlines compared to the fake ones. The predicted three-way interaction did not manifest. There was, however, an interaction between ground truth and match (see **Figure 3** in the *Figures* section). This indicates that participants show better discrimination between true vs. fake headlines when the political leaning of the stories is in line with the participant's ideology. This finding may potentially point to a relationship between alignment and accuracy in identifying real news stories.

Analysis 4:

We repeated analysis 3 now controlling for familiarity judgments. If participants use familiarity as a cue to the veracity of the story, we might expect the effect of ground truth to be reduced once we control for familiarity judgments. We used the following model:

$$Y = B_1 \times \text{Condition} + B_2 \times \text{Ground Truth} + B_3 \times \text{Match} + B_4 \times \text{Familiarity Judgement} + B_5 \times \text{Condition} \times \text{Match} + B_6 \times \text{Ground Truth} \times \text{Match} + B_7 \times \text{Condition} \times \text{Ground Truth} + B_8 \times \text{Condition} \times \text{Ground Truth} \times \text{Match} + \varepsilon$$

Unsurprisingly, familiarity is a good predictor for the veracity ratings, which indicates that people report a more accurate veracity judgement if they've seen the headline before. Even after controlling for familiarity judgments, however, ground truth continues to have a large effect on veracity judgments, with an estimate only slightly smaller than in analysis 3. The interaction of “ground truth” and “match” is also still significant, although smaller in size. These results potentially indicate that people are not fully relying on familiarity to form veracity judgments. The smaller interaction of match and ground truth, however, does indicate that familiarity may play a role in discriminating true from fake for stories that match a participants' politics. According to the beta coefficient, if an individual's beliefs line up with the leaning of the story (1 for match), and the story is true (1 for ground truth), then people judge the veracity to be slightly more true than for non-match true story, where there is no match in political affiliations. One could extrapolate an explanation from this. For example, perhaps people are more likely to be exposed to similar (but not the same) news that is true and in agreement with their beliefs. Likewise, they might not be exposed to significantly more fake news that aligns with their beliefs as compared to fake news in general, not in alignment with their belief.

5 DISCUSSION and CONCLUSION

There are a few key findings from our research. Firstly, people are generally good at discriminating between fake and real news. This performance could be for a number of reasons. For example, perhaps the true headlines are obviously true, and the false headlines are obviously false. Although we included the familiarity judgement to factor in whether or not participants have been exposed to that particular headline, this does not account for participants generally having knowledge about politics or being exposed to similar but not identical stories. Participants may have existing knowledge about related news that improved their accuracy in judging headlines. Alternatively, they may have made use of cues, such as the indicated source of the headline. In our study, the stimuli do include source. Generally speaking, the true headlines come from more mainstream outlets, and the fake headlines come from more obscure sources. Our study did not explore source as a predictor for veracity judgements, and so we did not randomize this element. However, it is possible that sources played a factor in discernment. Another potential confounder is that perhaps there is some trend in MTurk workers where they may simply consume more news compared to other people because their jobs are internet based, and thus would have generally more exposure to news stories, trends, and opinions. Regardless of political leaning, this could make stories easier to label, and could potentially account for how they were generally good at discerning true from fake news.

Although participants in the photo condition showed a small tendency to judge all headlines as more true than did participants in the control condition, the inclusion of the photo did not affect discrimination between true and fake headlines. Contrary to our prediction, the inclusion of a photo does not reduce discrimination or bring ratings of both true and fake

headlines toward “average”, where true things would be rated more false, and false things would be rated more true. A reason that this interaction is not significant could be related to the stimuli. Because the stimuli are from a few years ago, and some of the content in the stimuli were related to President Trump, who is no longer president, it is possible that the effects would be different if the stimuli were more current and topical to the contentious issues of today. In a previous pilot study with a sample size of $n=50$, this interaction between ground truth and condition was significant, but this effect did not replicate in the current larger preregistered study with a larger sample size of $n = 200$. This could be for many reasons, including that the pilot study results may have been spurious due to the small sample size. Additionally, we used a different participant recruitment strategy for this current study, specifically recruiting 100 democrats and 100 republicans, thus potentially changing the population compared to the pilot study.

Although not predicted, the study demonstrated an interesting interaction between ground truth and the matching variable. This indicates that people are generally better at discerning fake from real news when the headlines are in alignment with their political values. This could be because people are exposing themselves more to information that they agree with (in line with the confirmation bias theory) and thus have more knowledge about discerning fake from true news when it is related to their core beliefs. This could also indicate that participants are generally less familiar with headlines and news that do not align with their beliefs. It is worth noting that familiarity with the news story also significantly impacted their veracity judgements such that familiarity made people believe the story much more. Controlling for familiarity judgments decreased the magnitude of the condition x match interaction, suggesting that the interaction pattern is partly due to greater familiarity with headlines that match one’s political leaning.

Limitations of this research include using existing stimuli that are around four years old, because these topics are less relevant than current headlines. Additionally, the images we used were not originally included in these stimuli to specifically explore how images can impact fake news believability. It is possible that a different selection of images, perhaps ones that conveyed more information to the participants, would have had a stronger impact on believability. Additionally, although we asked participants not to look up the headlines on the internet, it is possible that they did so anyway, and there was no way to know whether or not they did this. It is also possible that they sought opinions from other people about the headlines before answering, which would also impact the results. Future research could use methodology that ensures individuals are solely working from their existing knowledge.

Research about the impact of various forms of media is very important to the future of fake news research. Fake news does not only come in the form of text, and fake news media outlets find innovative ways to deceive people. This includes video and photo deep fakes, misleading captions to videos, edited or manipulated visuals and audio, and the list goes on. Likewise, photo, video, audio, and text are all media that honest, reliable news outlets use to convey important information. In fighting this fake news epidemic, is it important to understand the cues that make people more susceptible to fake news in order to appropriately respond to fake news dissemination, and to be better informed on what forms of media increase believability of real news. Understanding what makes people believe news can better inform interventions such as social media flagging fake news as potentially misleading, or better censoring dangerously misleading information so that it is handicapped and thus less believable. Fake news is a massive problem in society and understand how and why it is believable is an important step in mitigating this problem.

6 FIGURES and TABLES

Table 1: Describing Data and Participants

	<i>Female</i>	<i>Age</i>	<i>White</i>	<i>Education</i>	<i>Search</i>	<i>Trust</i>
Minimum		0.00		0.00		1.00
1st Quartile		0.00		1.00		3.00
Median		1.00		1.00		4.00
Mean	0.47	0.51	0.87	0.93	1.99	3.57
3rd Quartile		1.00		1.00		4.00
Maximum		1.00		1.00		5.00
	<i>Party</i>	<i>Veracity Judg.</i>	<i>Familiarity</i>	<i>Democrat</i>	<i>Republican</i>	
Minimum		1.00	0.00			
1st Quartile		1.60	0.00			
Median		2.20	0.00			
Mean	1.47	2.20	0.17	0.53	0.47	
3rd Quartile		2.80	0.20			
Maximum		4.00	1.00			

Table 1: This table shows descriptive statistics for key variables. *Age*: a binary variable where 0 < 40 and 1 ≥ 40 years old. *Education*: A binary variable where 0 is less than college education and 1 is college education or greater. *Search* is whether or not participants search answers (binary 1,2). The vast majority did not. *Trust* is how much participants trust the media. The mean indicates that people generally trust the media. *Party*: a binary variable (1, 2) that demonstrates that approximately half the participants were democrats and vice versa. *Democrat* and *Republican* communicate the same sentiment.

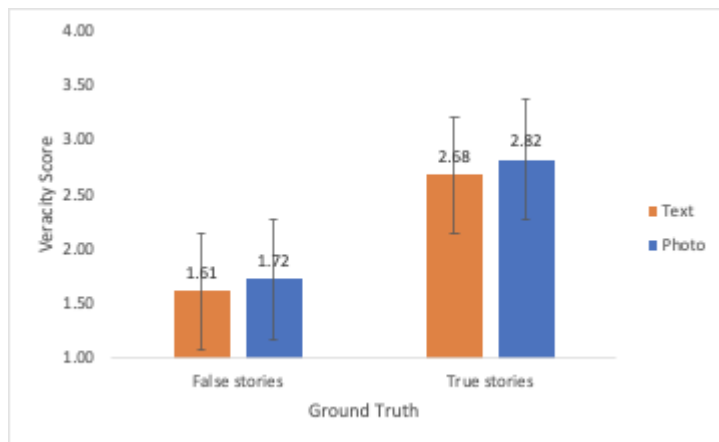
Table 2. Regressions

	(1)	(2)	(3)	(4)
β_0	1.608*	1.375*	1.607*	1.500*
	(0.04)	(0.153)	(0.044)	(0.040)
Condition	0.115*	0.106	0.130*	0.112
	(0.058)	(0.059)	(0.064)	(0.058)
Ground Truth	1.069*	1.069*	0.968*	0.846*
	(0.040)	(0.040)	(0.048)	(0.045)
Familiarity				1.175*
				(0.077)
Match			0.003	-0.0150
			(0.059)	(0.054)
Interactions				
Condition:GroundTruth	0.028	0.028	0.045	0.030
	(0.058)	(0.058)	(0.070)	(0.064)
Condition:Match			-0.044	-0.012
			(0.085)	(0.078)
GroundTruth:Match			0.303*	0.183*
			(0.084)	(0.077)
Condition:GroundTruth:Match			-0.052	-0.051
			(0.121)	(0.111)
Demographics				
Female		0.003		
		(0.053)		
Age		0.023		
		(0.022)		
Education		0.030		
		(0.032)		
White		0.041		
		(0.077)		

* Significant at $p < 0.05$

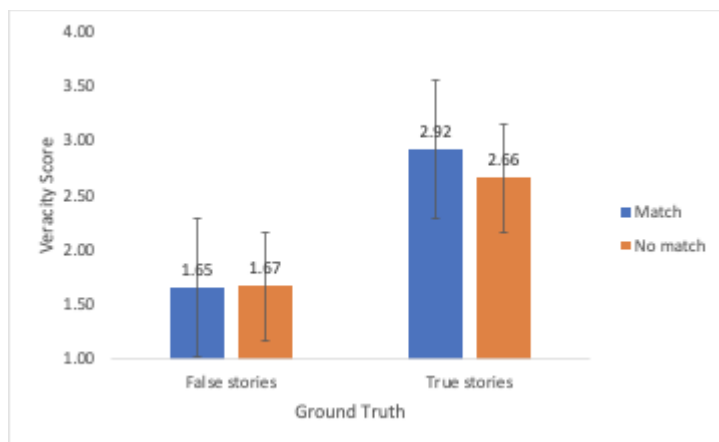
Table 2. This table represents the Beta Coefficients and significance of all four regressions. (1) Main regression. (2) Main regression including demographic variables as the control. (3) Main regression including matching variable and relevant interactions. (4) Main regression including matching variable and familiarity judgement . The sample size is $n = 200$.

Figure 2: The interaction between condition and ground truth



Description: This graph demonstrates that there is no significant difference in the responses based on condition. However, we do see that participants do rate false stories as more false and true stories as more true. This relationship is significant.

Figure 3: The interaction between match and ground truth



Description: This graph demonstrates that the interaction between match and ground truth is significant. When a story is false, whether or not the story politically aligns with the individual doesn't matter. However, when the story is true, stories that match the alignment of the participant are rated as higher than stories that do not politically align with the participant. The main effect of ground truth remains.

7 REFERENCES

- Leatherby, Lauren, et al. "How a Presidential Rally Turned Into a Capitol Rampage." *The New York Times*, The New York Times, 12 Jan. 2021, www.nytimes.com/interactive/2021/01/12/us/capitol-mob-timeline.html.
- Garimella et al, (2017). Image based Misinformation on WhatsApp. *Conference '17*
- Bago, B., Rand, D. G., & Pennycook, G. (2020). Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of Experimental Psychology: General*, 149(8), 1608–1613. <https://doi.org/10.1037/xge0000729>
- Bronstein et al (2019) Belief in Fake News is Associated with Delusionality, Dogmatism, Religious Fundamentalism, and Reduced Analytic Thinking. *Journal of Applied Research in Memory and Cognition*. <https://doi.org/10.1016/j.jarmac.2018.09.005>
- Clayton et al. (2019). Partisan motivated reasoning and misinformation in the media: Is news from ideologically uncongenial sources more suspicious?. *Japanese Journal of Political Science*. doi:10.1017/S1468109919000082
- Pennycook et al (2018). Prior Exposure Increases Perceived Accuracy of Fake news. *Journal of Experimental Psychology: General*. DOI: <http://dx.doi.org/10.1037/xge0000465>
- Pennycook and Rand (2019) Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality*. DOI: 10.1111/jopy.12476
- Pennycook and Rand (2018). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Elsevier*. <https://doi.org/10.1016/j.cognition.2018.06.011>

Sample Question



To the best of your knowledge, how accurate is the claim in the above headline?

Not at all accurate

☐

Not very accurate

☐

Somewhat accurate

☐

Very accurate

☐

Were you familiar with this news headline prior to today's study?

☐ Yes

☐ No