Overconfidence in news judgments is associated with false news susceptibility[†]

Benjamin A. Lyons
Department of Communication
University of Utah

Jacob Montgomery
Department of Political Science
Washington University in St. Louis

Andrew M. Guess
Department of Politics
Princeton University

Brendan Nyhan
Department of Government
Dartmouth College

Jason Reifler
Department of Politics
University of Exeter

Abstract

We examine the role of overconfidence in news judgment using two large nationally representative survey samples. First, we show that three in four Americans overestimate their relative ability to distinguish between legitimate and false news headlines; respondents place themselves 22 percentiles higher than warranted on average. This overconfidence is in turn correlated with consequential differences in real-world beliefs and behavior. We show that overconfident individuals are more likely to visit untrustworthy websites in behavioral data; to fail to successfully distinguish between true and false claims about current events in survey questions; and report greater willingness to like or share false content on social media, especially when it is politically congenial. In all, these results paint a worrying picture: the individuals who are least equipped to identify false news content are also the least aware of their own limitations and therefore more susceptible to believing it and spreading it further.

Concern about public susceptibility to false news is widespread. However, though Americans believe confusion caused by false news is extensive, relatively few indicate having seen or shared it (Barthel, Mitchell and Holcomb, 2016) — a discrepancy which suggests that members of the public may not only have a hard time identifying false news, but fail to recognize their own deficiencies at doing so (Davison, 1983; Sun, Shen and Pan, 2008; Kruger and Dunning, 1999; Dunning, 2011). Such overconfidence may make individuals more likely to inadvertently expose themselves to misinformation and to participate in its spread. If people incorrectly see themselves as highly skilled at identifying false news, they may unwittingly be more likely to consume, believe, and share it, especially if it conforms to their worldview.

Overconfidence plays a key role in shaping behavior, at least in some domains (e.g., Ehrlinger et al., 2008; Ferraro, 2010; Ortoleva and Snowberg, 2015b; Sheffer and Loewen, 2019; Kovacs, Lagarde and Cairns, 2020). However, we know very little about its potential role in the spread of false news. Even basic descriptive data on the phenomenon of overconfidence in news discernment (the ability to distinguish false from legitimate news) is yet to be established. How pervasive is overconfidence? Is overconfidence related to false news exposure? Are overconfident individuals actually more likely to hold misperceptions or share false stories? We currently lack answers to these questions.

In this paper, we examine the relationship between perceived and actual ability to distinguish between false and legitimate information, drawing on a theoretical framework for understanding biased self-perception (Kruger and Dunning, 1999). In two large nationally representative samples ($N = 8,285$), respondents completed a novel discernment task evaluating the accuracy of a series of headlines as they appear on Facebook. They were further asked to rate their own abilities in discerning false news content relative to others. We use these two measures to assess overconfidence among respondents and how it is related to beliefs and behaviors.

Our results paint a worrying picture. The vast majority of respondents (about 90%) reported that they are above average in their ability to discern false and legitimate news headlines, and many Americans substantially overestimate their abilities. Accordingly, people's self-perceptions are only weakly correlated with actual performance. Further, using data measuring respondents' online behavior, we show that those who overrate their ability more frequently visit websites known to spread false or misleading news. These overconfident respondents are also less able to distinguish between true and false claims about current events and report higher willingness to share false content, especially when it aligns with respondents' political predispositions. Although discernment ability is a strong predictor of these outcomes, an alternative analysis using a "residualized" measure of overconfidence net of actual ability also explains additional variance in these behaviors.

In the next section, we review existing research on overconfidence and how we expect it to operate for news discernment. Our methods section describes our research design, including a novel task assessing respondents' news discernment abilities. Our results show that overconfidence is both common and associated with a range of undesirable media-related behaviors. Although our design does not allow us to identify the causal effect of overconfidence, these findings suggest that the mismatch between one's perceived ability to spot false stories and people's actual abilities may play an important and previously unrecognized role in the spread of false information online.

Who spreads false news?

Which individuals are more likely to engage with, believe, and spread dubious news? One body of research emphasizes the role of partisan predispositions or motivated reasoning in the assessment of news content (Vegetti and Mancosu, 2020) and exposure to it and sharing of it (Grinberg et al., 2019; Guess, Nyhan and Reifler, 2020). A second literature considers how improving individuals' information evaluation and digital literacy skills can reduce their vulnerability to false information online (Guess, Lerner, Lyons, Montgomery, Nyhan, Reifler and Sircar, 2020; Roozenbeek and van der Linden, 2019). Finally, other studies focus on the role of purposeful reasoning processes in reducing individual vulnerability to misinformation. People who think more analytically or are more deliberative in their evaluation of news claims rate false news as less accurate (Pennycook and Rand, 2018*a*; Bago, Rand and Pennycook, 2020). Conversely, people who tend to rely on emotion as they process information or wrongly claim familiarity with nonexistent entities are more likely to see false headlines as accurate (Pennycook and Rand, 2018*b*; Martel, Pennycook and Rand, 2020).

Our research builds on cognitive style accounts by examining the disparity between people's ability to spot false news and their beliefs about their skill in doing so. This approach is intended to assess the contribution of cognition as well as metacognition to engagement behaviors. As we argue below, overconfidence in one's ability to distinguish between legitimate and false news may help account for whether and how individuals engage with false or dubious online content (e.g., liking or sharing). To put the point more directly, some portion of the public is likely to be especially vulnerable to false information precisely because they do not realize that they are, in fact, vulnerable to false information. As a result, these individuals may be more likely to unknowingly consume, believe, and share false news.

The Dunning-Kruger effect for news discernment

Building on prior studies of perceptual bias in self-assessments, we test for a Dunning-Kruger effect (DKE) in false news discernment. The DKE describes a general tendency of poor performers in social and intellectual domains to be unaware of their own deficiency (Kruger and Dunning, 1999). By contrast, the most competent performers slightly underestimate their own ability relative to others due to a form of false consensus effect in which they assume others are performing more similarly to themselves than they really are (Ross, Greene and House, 1977). This pattern arises whether researchers elicit comparative self-evaluations (ratings of performance relative to peers) or self-evaluations using absolute scales (Dunning, 2011).

DKE research contends that poor performers suffer from a double bind: not only does a lack of expertise produce errors in the first place, it also prevents recognition of these errors and awareness of others' capabilities. In studies of perception and performance, people in the bottom quartile of performers have tended to provide the most upwardly distorted self-perceptions. For instance, Anson (Anson, 2018) finds that individuals who perform worst on a quiz measuring basic political knowledge rate their own performance the same or even better than high performers.

The reported overconfidence of underperformers is not erased by financial or social incentives (Ehrlinger et al., 2008) and is corroborated by real-world behavior (e.g., in (not) selecting insurance for exam performance (Ferraro, 2010)). These studies suggest that low performers genuinely believe in their own abilities and are not simply making face-saving expressions of self-worth. Further, past research shows that overconfidence is more common when people have reason to see themselves

as knowledgeable or competent — i.e., if the subject is not arcane and is prevalent in everyday life (Dunning, 2011). Given its familiarity, judgments of news accuracy are likely to fit the DKE pattern as do knowledge about either politics (Anson, 2018) or vaccines (Motta, Callaghan and Sylvester, 2018). We therefore propose the following research question:[1]

> *Research Question 1: To what extent will people who are least accurate at distinguishing between legitimate and false news overrate their ability to distinguish mainstream from false news?*

Importantly, the DKE predicts that low performers will not recognize how poorly they performed in *relative terms*, not that low performers will think they perform best. We therefore do not expect that low performers will think they are the best at our task of distinguishing between legitimate and false news. Instead, we will examine the extent to which poor performers do not recognize that they are worse than most others at the task.

Does overconfidence matter?

Importantly, the DKE may have downstream effects on behavior. Because overconfident individuals fail to recognize their own poor performance, they are less able to improve their domain-specific skills. For instance, several studies find that overconfident individuals learn the least in classroom settings (Pazicni and Bauer, 2014).[2] We therefore expect that overconfidence in news discernment will be associated with a variety of tendencies including exposure to false news, belief in its accuracy, and sharing it with others.

To begin, we expect a positive association between overconfidence and visits to false news websites. The DKE implies less ability to discern which news stories are false when an individual is exposed (e.g., on a social media platform), combined with lesser awareness of this discernment deficiency, which would lead to greater incidental exposure to false news stories. Similarly, overconfidence may be seen as a form of invulnerability bias in which assumed mastery leads people to feel little need to take preventative actions (e.g., to be cautious or engage in deliberate thinking about which sites one visits), which may produce additional exposure to questionable media messages (Douglas and Sutton, 2004; Hansen et al., 2010). We therefore propose the following research question:

> *Research Question 2: Is overconfidence in one's ability to distinguish mainstream from false news positively related to false news exposure?*

In addition, overconfidence may make people less likely to question a dubious news story's veracity, as high confidence is associated with less reflection (Thompson, Turner and Pennycook, 2011; Pennycook et al., 2017). As a result, people who are overconfident may be more willing to

---

[1]We filed a preregistration for this project prior to accessing the data. We report a "populated pre-analysis plan" (Duflo et al., 2020) that details our preregistered hypotheses and analysis plan and identifies which main text findings are preregistered in SI Appendix E.

[2]A related literature details the confidence with which individuals hold political misperceptions (Pasek, Sood and Krosnick, 2015; Li and Wagner, 2020; Graham, 2020). This works shows that many people are somewhat aware of their ignorance and therefore many misperceptions are not confidently held (Graham, 2020), and these individuals are more likely than the confidently wrong to update their beliefs in response to corrections (Li and Wagner, 2020).

accept false claims and to engage with false content in the form of liking or sharing these stories (similarly, recent work suggests people generally lack awareness of their susceptibility to inaccurate "general knowledge" claims they come across when reading works of fiction (Salovich and Rapp, 2020)). Further, previous research indicates that individuals are generally more likely to believe false claims when they are consistent with their own prior political beliefs (Flynn, Nyhan and Reifler, 2017). Therefore, we would expect that the relationship between overconfidence and beliefs and engagement will be strongest when the content involved aligns with respondents' partisan preferences.

> *Research Question 3: (a) Is overconfidence positively related to holding misperceptions on specific topics? (b) Is this relationship stronger when the claim is politically congenial?*

> *Research Question 4: (a) Is overconfidence positively related to self-reported willingness to like or share false content? (b) Is this relationship stronger when the claim is politically congenial?*

Results

We first describe the DKE in our data. We divide the sample into four quartiles based on respondents' *actual* performance in our discernment task. For each of these four groups, we calculate the mean score for both actual and perceived ability (percentiles ranging from 1–100), which we present in Figure 1. As expected, actual performance closely tracks the idealized 45-degree line when we plot the mean performance score in each quartile. However, for perceived ability, we see a much flatter line. Perceived ability increases modestly across our measure of actual ability. The mean self-reported percentile for individuals in the bottom quartile in actual ability (i.e., the 1st–25th percentile) is 63 in the Oct./Nov. survey and 64 in the Nov./Dec. survey. This quantity rises to only 74 for the top quartile in both surveys. In other words, those who are in the bottom quartile in actual performance rate themselves as being in about the 63rd/64th percentile, a vast overestimate of their own performance. While those in the top quartile of actual performance rate their perceived ability higher than those in the bottom quartile do, they underestimate where they rank in actual ability.[3]

In general, performance is only weakly associated with perceived ability (Oct./Nov. $r = .08$; Nov./Dec. $r = .10$), as shown in SI Appendix Figure B3. Moreover, average self-reported percentile (69th) is well above 50 (one-sample t-test $p < .005$), indicating that many people are overconfident. As Figure 1 illustrates, this overconfidence is concentrated most heavily among individuals in the

---

[3]Critics of DKE analyses like the one presented in Figure 1 argue that it does not reflect the proposed mechanism — metacognitive differences (i.e., perception accuracy) between high and low performers — and is instead the result of systematic bias or measurement error; e.g., regression to the mean and the better-than-average effect (Burson, Larrick and Klayman, 2006; Krueger and Mueller, 2002). Proposed alternative accounts of the DKE have led to vigorous theoretical and empirical debates (Dunning, 2011; Ehrlinger et al., 2008; Kruger and Dunning, 2002; Feld, Sauermann and De Grip, 2017; Schlösser et al., 2013; Gignac and Zajenkowski, 2020; Miller et al., 2019; McIntosh et al., 2019). Though no consensus has emerged, recent work suggests that metacognitive differences, general biases in self-estimation, and statistical artefacts each contribute to the DKE (McIntosh et al., 2019).

bottom quartile. That is, the individuals whose performance is objectively at the lowest level are the most *overconfident* in their abilities.

In line with prior work, male respondents display more overconfidence (Niederle and Vesterlund, 2007; Ortoleva and Snowberg, 2015*b*; Ferraro, 2010), and overconfidence is negatively associated with general political knowledge. There is no association with age (Prims and Moore, 2017) despite age-based disparities in exposure to false news (Grinberg et al., 2019; Guess, Nyhan and Reifler, 2020). Finally, Republicans are more overconfident than Democrats (Ortoleva and Snowberg, 2015*a*), which is not surprising given the lower levels of media trust they report (see SI Appendix C, which shows that mass media trust and media affect are both negatively associated with overconfidence). We report pre-registered analyses regarding demographics in greater depth in a separate manuscript (see SI Appendix E for details).

False news exposure

We next examine whether visits to false news websites are associated with overconfidence. Building on prior research examining the difference between subjective self-perceptions and objective performance (Ortoleva and Snowberg, 2015*b*; Kovacs, Lagarde and Cairns, 2020; Bregu, 2020; Lambert, Bessière and N'Goala, 2012; Sheffer and Loewen, 2019), we measure this concept as the difference between self-reported relative performance and our objective measure of relative performance. As Parker and Stone (Parker and Stone, 2014) argue, the difference score measure we employ here is appropriate when the theoretical mechanism of interest is overconfidence rather than self-assessed ability *per se* (i.e., controlling for ability). We are interested in the miscalibration between these components because the DKE relies on the *double bind* of low ability paired with a lack of awareness.

We estimate ordinary least squares models using binary measures of false and mainstream news exposure for both surveys (Oct./Nov. and Nov./Dec.). In each of these models, which are estimated using survey weights, we include a set of standard covariates as well as a measure of the ideological orientation of respondents' news diet. Finally, we re-scale our measure of overconfidence to range from -1 to 1 rather than from -100 to 100 to aid in interpretation. Results are shown in Table 1 and Figure 2. The baseline exposure rate to false news in the Oct./Nov. survey was 6.5%. We find that overconfidence is associated with greater rates of exposure in that survey ($\beta = .06, SE = .02, p < .01$). Specifically, respondents at the 95th percentile of overconfidence were about six percentage points more likely to have been exposed to false news in the post-survey period than the those at the 5th percentile conditional on demographics. Similarly, those at the maximum value of overconfidence were about 11 percentage points more likely to have been exposed than those at the minimum. The relationship is not statistically significant for the Nov./Dec. model, but the sample size for that survey is significantly reduced ($n = 767$). When we instead pool the data, the results are nearly identical to the results for the Oct./Nov. survey (see SI Appendix Table F1 for results from logit models, which are substantively identical). One concern is that overconfident individuals may simply be more (or less) likely to visit online news websites in general. To test for this, we also estimate identical regressions with mainstream news exposure as the dependent variable. We find that overconfidence is not associated with our binary measure of mainstream news exposure after accounting for demographics.

5

Table 1: Overconfidence and news exposure (binary measures)

| | Oct./Nov. | | Nov./Dec. | | Pooled | |
	False	Mainstream	False	Mainstream	False	Mainstream
Overconfidence	0.0609**	-0.0450	0.0003	-0.0007	0.0569***	-0.0415
	(0.0231)	(0.0505)	(0.0003)	(0.0006)	(0.0186)	(0.0411)
Constant	-0.0815*	0.4645***	-0.0225	0.3735***	-0.0715*	0.4419***
	(0.0354)	(0.1010)	(0.0498)	(0.1199)	(0.0298)	(0.0799)
Control variables	✓	✓	✓	✓	✓	✓
R^2	0.17	0.11	0.08	0.17	0.11	0.11
N	1780	1780	767	767	2547	2547

* $p < .05$, ** $p < .01$, *** $p < .005$ (two-sided). Cell entries are OLS coefficients estimated using survey weights. The overconfidence measure subtracts the respondent's actual percentile from their self-rated percentile and is re-scaled to range from -1 to 1. False news exposure is coded as 1 if the respondent visited any such domain and 0 otherwise. Mainstream news exposure is coded as 1 if the respondent visited any such domain and 0 otherwise. All models include controls for Democrat, Republican, college education, gender, nonwhite racial background, age, and media diet slant.

Topical misperceptions

Next, we examine the association between overconfidence and ability to distinguish between true and false claims about political events that were topical at the time the surveys were fielded. Here we examine a misperceptions battery from the October/November survey measuring beliefs in claims related to Brett Kavanaugh's Supreme Court nomination. These regression models are again estimated using survey weights and our set of standard covariates. We also again re-scale our measure of overconfidence to range from -1 to 1 rather than from -100 to 100 to aid in interpretation. The main results of this analysis are shown in Table 2. The first column shows the results for the two false statements provided to respondents. We include fixed effects for each statement to account for their differing baseline levels of plausibility and cluster at the respondent level to account for correlations between their ratings across headlines. We find no main effect of overconfidence on belief in these false claims in isolation (and no evidence that this relationship is moderated by congeniality). The second column, however, shows results for difference scores (discernment), which subtracts the perceived accuracy of false claims from that of true claims. Higher discernment scores reflect greater belief in true statements relative to false ones. The negative coefficient ($\beta = -.37, SE = .06, p < .005$) thus indicates that overconfidence is negatively associated with discernment ability on these topical claims.

Self-reported engagement

We next turn to our measure of self-reported engagement intention (intent to like or share a post on social media). These regression models again include a set of standard covariates, and we re-scale our measure of overconfidence to range from -1 to 1 rather than from -100 to 100, to aid in interpretation. We use survey weights in all models. These results are shown in Table 3. The first and third columns show results at the headline level, where the outcome is a four-point scale of intention to either share or like a false story. Overconfidence has a clear positive relationship with

Table 2: Overconfidence and topical misperceptions

	False	Difference score
Overconfidence	0.1123	-0.3667***
	(0.0966)	(0.0604)
Congeniality	0.8557***	
	(0.0475)	
Overconfidence × congeniality	-0.0390	
	(0.1296)	
Constant	1.9550***	-0.2223*
	(0.1056)	(0.0932)
Control variables	✓	✓
Statement fixed effects	✓	
R^2	0.16	0.15
N (statement)	4872	
N (respondent)	2444	2904

* $p < .05$, ** $p < .01$, *** $p < .005$ (two-sided). Cell entries are OLS coefficients. Respondents rated the accuracy of four statements regarding the Kavanaugh appointment on four-point scales. The first model's outcome variable is perceived accuracy of false statements only. The second model's outcome variable is the difference in the mean perceived accuracy of true and false statements. The overconfidence measure subtracts the respondent's actual percentile from their self-rated percentile, and is re-scaled to range from -1 to 1. Controls: Democrat, Republican, college education, gender, nonwhite racial background, and age.

liking or sharing false stories (RQ4a). Moreover, this relationship varies as a function of partisan congeniality (RQ4b). Likewise, the second and fourth columns of Table 3 use the average difference in engagement intention across true and false headlines as a measure of discernment. The results show that overconfidence is negatively related to discernment in which stories respondents would engage with. Overconfident individuals are thus not merely more likely to engage with news content in general, but instead are specifically more inclined to share false stories versus mainstream ones relative to respondents who are less overconfident.

Alternative specifications

We construct our primary independent variable above as: Overconfidence = (Perceived ability - Actual ability). We view this measurement strategy as appropriate for two reasons. First, this approach is consistent with how overconfidence has been measured in related studies of its behavioral effects (Ortoleva and Snowberg, 2015*b*; Sheffer and Loewen, 2019; Kovacs, Lagarde and Cairns, 2020; Bregu, 2020; Lambert, Bessière and N'Goala, 2012). Second, and more importantly, our theory is explicitly about the *difference* between perceived and actual ability and not about the independent role of either component. Thus, the main regressions of interest are (broadly) structured as:

$$y_i = \gamma_0 + \gamma_1 \left(\text{Perceived ability} - \text{Actual ability} \right) + \varepsilon_i, \tag{1}$$

where y_i represents the outcome of interest and ε_i is our error term.

We use this specification because we have no theoretical expectations about the independent role of these predictors — our theory is about the mismatch between actual and perceived ability (Parker

Table 3: Overconfidence and engagement intention

	Oct./Nov.		Nov./Dec.	
	False	Diff. score	False	Diff. score
Overconfidence	0.6690***	-0.4051***	0.7610***	-0.3984***
	(0.0691)	(0.0193)	(0.0466)	(0.0283)
Congeniality	0.1893***		0.1902***	
	(0.0224)		(0.0151)	
Overconf $\times$ congenial	0.2394***		0.2692***	
	(0.0721)		(0.0511)	
Constant	1.0620***	0.0520	1.1020***	0.0600
	(0.0242)	(0.0270)	(0.0139)	(0.0369)
Control variables	✓	✓	✓	✓
Headline fixed effects	✓		✓	
R^2	0.11	0.15	0.13	0.13
N (headline)	10194		14720	
N (respondent)	2549	2566	3680	3717

* $p < .05$, ** $p < .01$, *** $p < .005$ (two-sided). Cell entries are OLS coefficients. DVs are based on self-reported intention to "like" or "share" each of the articles in the headline task on Facebook (four-point scales; 1 = not at all likely, 4 = very likely). These questions are asked only of respondents who report using Facebook. The first model's outcome variable is engagement intent for false headlines only. The second model's outcome variable is the difference in the mean engagement intent for mainstream and false headlines. The overconfidence measure subtracts the respondent's actual percentile from their self-rated percentile, and is re-scaled to range from -1 to 1. Controls: Democrat, Republican, college education, gender, nonwhite racial background, and age.

and Stone, 2014). However, it is still worthwhile to consider whether one of these factors (perceived or actual ability) is responsible for our results. In particular, it is important to try to isolate the effects of perceived ability given prior findings showing that false news belief and exposure are related to individual-level differences in analytical thinking skills and reasoning ability (Pennycook and Rand, 2018*a*; Bago, Rand and Pennycook, 2020; Guess, Lerner, Lyons, Montgomery, Nyhan, Reifler and Sircar, 2020; Guess, Nyhan and Reifler, 2020).

We therefore consider two alternative model specifications that seek to estimate the direct association between *perceived* ability and our outcome measures independent of its relationship to *actual* ability below. First, we attempt to "residualize" perceived ability, an approach that has become standard in personality and social psychology Belmi et al. (2020). Second, we disaggregate the two components and include them independently in a regression. (These approaches are mathematically quite similar, but we include them both for the sake of completeness).[4] To account for the fact that these three approaches each have unique weaknesses, recent work has suggested all three be employed Belmi et al. (2020).

Residualizing perceived ability

We begin by following the strategy outlined in Anderson et al. (Anderson et al., 2012), which uses a residualized measure of perceived ability. Specifically, we first fit the regression

$$\text{Perceived}_i = \text{Actual}_i + \delta_i \tag{2}$$

where δ_i is the residual error term. Assuming this model is correct, we can then use the estimated residual error $\hat{\delta}_i$ as a measure of perceived ability that is unrelated to actual ability. We then fit a model such as

$$y_i = \beta_0 + \beta_1 \hat{\delta}_i + \varepsilon_i, \tag{3}$$

where β_1 is intended to represents the independent relationship between (residualized) perceived ability and the outcome.

With this residualization approach, we start with news exposure in SI Appendix Table F2. We find a positive correlation between residualized perceived ability and exposure but it is only statistically significant for the pooled model (Oct./Nov.: $\beta = 0.07$, $p > .05$; Nov./Dec.: $\beta = 0.02$, $p > .05$; Pooled: $\beta = 0.06$, $p < .05$). This result differs from our primary analysis only in that coefficient for the Oct./Nov. survey is not statistically significant, though as in the primary analysis, it is similar to the pooled coefficient. Turning to topical misperceptions, SI Appendix Table F3 shows that there is a significant interaction between residualized perceived ability and congeniality ($\beta = 0.59$, $p < .05$), indicating that overconfident individuals are more likely to believe in false statements that are consistent with their prior beliefs. This result is more favorable for our theory than the one reported in the primary analysis. However, unlike the primary results, residualized perceived ability is not significantly associated with decreased discernment between true and false claims. Finally, SI Appendix Table F4 shows that residualized perceived ability is positively associated with liking or sharing false stories (Oct./Nov.: $\beta = 0.33$, $p < 0.01$; Nov./Dec.: $\beta = 0.47$, $p < .005$). These relationships are strongest for congenial stories (Oct./Nov.: $\beta = 0.41$, p<.005; Nov./Dec.: $\beta = 0.47$,

---

[4]Indeed, the results would be identical if we used the full set of covariates from the disaggregated regression in our residualization process.

$p < .005$). However, there is again no significant association with discernment between mainstream and false news in either wave (Oct./Nov.: $\beta = -0.10$, $p > .05$; Nov./Dec.:$\beta = 0.03$, $p > .05$).

Disaggregating overconfidence

Our second approach is to include perceived and actual ability as two separate independent variables in our model. Though our theory focuses on overconfidence, one might expect the coefficient for perceived ability to be positive and the coefficient for actual ability to be negative for the outcome measures we consider. To illustrate this idea, we simulate data according to formulas that assume a data-generating process in which overconfidence is linearly associated with some outcome measure per our theory (see SI Appendix Table F5). These results, which show perceived ability is positively associated with the outcome and actual ability is negatively related to the outcome, suggest that the disaggregation approach will provide the correct conclusion.

However, it is important to emphasize a few important limitations before presenting our disaggregated results. First, this approach assumes that the component measures do not affect one another despite the fact that self-perception and performance likely do so (Belmi et al., 2020; Dunning, 2011). Second, this strategy is more difficult to interpret because both coefficients relate to the theory of interest. Increased perceived ability (controlling for actual ability) is an indicator for overconfidence but so too is decreasing actual ability (controlling for perceived ability). Interpreting either coefficient in isolation with respect to our theory is therefore difficult, especially in more complex models (i.e., those that include interaction terms; see Parker and Stone (Parker and Stone, 2014) for more extensive discussion of this point). Third, the simulated results are based on the assumption of constant levels of measurement error between the perceived and actual ability variables. If measurement error varies between them, however, it may *appear* as if only one of the two variables is important despite the fact that both are equally weighted in the true data generating process. To illustrate this point, we conduct a version of the simulation described above but now add additional measurement error to the *observed* perceived ability variable included in the disaggregated regression.[5] We thus assume the *same* data generating process where *overconfidence* drives our results but now add differential measurement error for perceived ability. The results in SI Appendix Table F6 now show a null result for perceived ability and a significant negative association with actual ability. Researchers who failed to consider the possibility of differential measurement error might mistakenly infer that it is only actual ability that drives these results. This scenario seems empirically plausible. *A priori* we would not expect equal rates of measurement error between these components. Specifically, actual ability is measured via a series of twelve objective evaluation tasks that are combined into an aggregate score. By contrast, perceived ability is measured as the average of two self-assessment survey items. Standard psychometric theory would suggest higher rates of measurement error for the perceived ability indicator.

With these caveats, we turn to our disaggregated results below. First we re-examine RQ1, which predicts that overconfidence will be related to differential rates of exposure to false news websites. The disaggregated models are shown in SI Appendix Table F7. Consistent with the extrapolation from our theory described above, the perceived and actual ability coefficients are signed in opposite directions, but only the actual ability coefficients are significant for the Oct./Nov. ($\beta = -0.06$, p<.05) and pooled samples ($\beta = -0.06$, p<.05) (perceived ability: Oct./Nov. $\beta = 0.06$, pooled

---

[5]Specifically, we add normal noise with standard deviation 2.5.

$\beta = 0.05$, both not significant). These results suggest either that actual ability is more important than perceived ability or is measured with less error per our discussion above. As in the primary analysis, neither is significant for the Nov./Dec. sample. Next, we turn to the topical misperceptions results (SI Appendix Table F8). In the primary analysis, we find no main effects or interactions with congeniality but do find a main effect for the *difference* outcome. When we disaggregate, we do find main effects for the actual ability measure in both analyses. The interaction terms, however, tell a complicated story. There is a positive significant interaction between perceived ability and congeniality ($\beta = 0.62$, $p < .01$), indicating that more overconfident individuals are more likely to believe false claims that are congenial to their prior beliefs. However, there is also a positive significant coefficient for the interaction with actual ability ($\beta = 0.26$, $p < .05$), which suggests that overconfidence (decreased actual ability controlling for perceived ability) increases belief in false stories only when they are not congenial. Finally, we turn to the results for engagement intentions (SI Appendix Table F9). For the headline-level analyses, the results again mirror the findings in the primary analysis with both the perceived ability and actual ability coefficients being significant (but signed in opposite directions). The interactions with headline congeniality are also both significant and correctly signed. For the difference score analysis, the results are more mixed. The actual ability coefficient is significant and positive for the Oct./Nov. sample ($\beta = 0.60$, $p < .005$) and the Nov./Dec. sample ($\beta = 0.61$, $p < .005$). However, the perceived ability coefficient is not significant for the Oct./Nov. sample ($\beta = 0.03$, $p > .05$) and significant but incorrectly signed for the Nov./Dec. analysis ($\beta = 0.19$, $p < .005$).

In all, these additional tests provide a somewhat mixed picture. While our results certainly show that not all results are driven purely by the actual ability measure, some of the evidence suggests that actual ability could be playing a crucial role for some of our results. However, we cannot rule out the possibility that these differences are attributable to differential measurement error. In several cases, the point estimates for perceived and actual ability are quite similar and the main differences in our inferences are the result of our estimates of perceived ability being more imprecise (e.g., SI Appendix Table F7).

Discussion

We find that respondents tend to think they are better than the average person at news discernment, and perceived ability is only weakly associated with actual ability, with the worst performers also being the most overconfident. Importantly, overconfidence is associated with a range of normatively troubling outcomes, including visits to false news websites in online behavior data. The overconfident also express greater willingness to share false headlines and are less able to discern between true and false statements about contemporaneous news events. Notably, the overconfident are particularly susceptible to congenial false news. These results suggest that overconfidence may be a crucial factor for explaining how false and low-quality information spreads via social media.[6] Many people are simply unaware of their own vulnerability to misinformation. Targeting these overconfident individuals could be an important step toward reducing misinformation on social media sites, though how best to do so remains an open question. Other research finds that the behavioral

---

[6]SI Appendix C also explores whether and how overconfidence is related to trust in the media. We show that overconfidence is negatively associated with trust in the mainstream media, but positively associated with trust in information seen on Facebook.

effects of high confidence and weak performance include resistance to help, training, and corrections (Dunning, 2011; Li and Wagner, 2020; Sheldon, Dunning and Ames, 2014). An incorrect view of one's ability to detect false news might reduce the influence of new information about how to assess media items' credibility as well as willingness to engage with digital literacy programs. For this reason, it may be important to better understand the roots of overconfidence, from demographics (Mondak and Anderson, 2004) to domain involvement (Perloff, 1989) to social incentives (Cheng et al., 2020; Anderson et al., 2012), and how they apply in the case of perceptions of news discernment.

These results should also be understood in the context of their limitations. Most critically, our analyses are correlational and thus face concerns about endogeneity. In this context, we have specific *ex ante* reasons to suspect that the relationship between overconfidence and our outcome measures is at least partially endogenous. For instance, habitual exposure to false news might lead to poorly calibrated estimates of one's ability to detect it, especially given the tendency for people to treat incoming information as true and the subsequent effects this can have on feelings of fluency (Brashier and Marsh, 2020). Overconfidence and false news engagement could even mutually reinforce one another over time (Slater, 2007). Future work must determine the extent to which overconfidence plays a causal role in the behaviors with which we show it is associated. One possible direction would be to experimentally manipulate overconfidence by informing respondents about their relative performance. Another approach might be to manipulate individuals' self-perception by randomly assigning them a competency score, although such a study would require careful ethical consideration.

Beyond issues of endogeneity, it is important to carefully interpret the associations we detect. Based on prior work regarding the role of purposeful reasoning (Pennycook and Rand, 2018*a*; Bago, Rand and Pennycook, 2020) and literacy skills (Guess, Lerner, Lyons, Montgomery, Nyhan, Reifler and Sircar, 2020) in individual vulnerability to misinformation, we would assume discernment ability itself — from which our overconfidence measure is in part derived — drives engagement with this content. Unsurprisingly, we find that people who are worse at discerning between legitimate and false news in the context of a survey are worse at doing so in their browsing habits. Further, actual ability is a stronger predictor than perceived ability, though the effect sizes are similar; as noted, this discrepancy may be a reflection of greater measurement error in our measure of perceived ability. However, our results also show that *inflated perceptions of ability* are independently associated with engaging with misinformation, suggesting perceived ability net of actual ability may be a further source of vulnerability (i.e., an additional, metacognitive component). Specifically, when residualized, perceived ability net of actual ability is associated with dubious news site exposure, misperceptions, and sharing intent. It is not our goal here to argue that overconfidence supersedes ability itself as the key predictor or cause of vulnerability to misinformation, nor do our findings support this interpretation. Indeed, our results lend further support to work that shows ability deficits are a serious issue in this domain. Further, because excess confidence is associated with less reflection (Pennycook et al., 2017; Thompson, Turner and Pennycook, 2011), the ways that discernment ability and overconfidence influence engagement with dubious information may be linked. Ultimately, adjudicating between these accounts would require further improvements to the measurement of overconfidence, which remains a complicated endeavor in all research contexts (Parker and Stone, 2014; Dunning, 2011; Belmi et al., 2020). We rely on overconfidence as measured by the difference between actual and self-assessed performance on a news discernment task. Future research should explore different approaches to measuring overconfidence in this domain

and assess how they relate to who views, believes, and spreads false news content. In particular, scholars should consider how to measure perceived ability with more precision and/or seek to directly manipulate these concepts in isolation to understand their independent effects.

Finally, though we replicate our results in multiple samples, further efforts to demonstrate that the relationship we observe holds in other contexts would be valuable. First, work should validate these results with mobile data and with data that allow us to observe actual sharing behavior in addition to self-reported sharing (though they appear to correspond at least to some extent (Mosleh, Pennycook and Rand, 2020)). Likewise, our data come from the American context, though based on cross-national findings regarding the pervasive nature of overconfidence (Stankov and Lee, 2014), it is reasonable to believe the outcomes are not unique to the U.S. and may be even more worrisome elsewhere.

Ultimately, our results provide new evidence of an important potential mechanism by which people may fall victim to misinformation and disseminate it online using survey and behavioral data from multiple large national samples. Understanding overconfidence may be an important step toward better understanding the public's vulnerability to false news and the steps we should take to address it.

Materials and methods

To answer our research questions, we draw on data from two novel two-wave survey panels conducted by the survey company YouGov during and after the 2018 U.S. midterm elections, allowing us to replicate our analyses across time and samples:

- A two-wave panel study fielded October 19–26 (wave 1; N = 3,378) and October 30–November 6, 2018 (wave 2; N = 2,948)

- A two-wave panel study fielded November 20–December 27, 2018 (wave 1; N = 4,907) and December 14, 2018–January 3, 2019 (wave 2; N = 4,283)

Respondents were selected by YouGov's matching and weighting algorithm to approximate the demographic and political attributes of the U.S. population (see SI Appendix A). Participants were ineligible to take part in more than one study. Both surveys in this research were approved by the institutional review boards of the University of Exeter, the University of Michigan, Princeton University, and Washington University in St. Louis. All subjects gave informed consent to participate in each survey. The pre-analysis plans are available at `https://osf.io/fr4k5` and `https://osf.io/r2jvb`.[7]

Measuring discernment ability: News headline rating task

In each survey, we asked respondents to evaluate the accuracy of a number of headlines on a four-point scale ranging from "Not at all accurate" (1) to "Very accurate" (4). The articles, all of which

---

[7]Participants received an orthogonal treatment related to media literacy in both surveys. Due to a programming error, all respondents received the treatment in the Oct./Nov. survey. The results of this study are reported in (Guess, Lerner, Lyons, Montgomery, Nyhan, Reifler and Sircar, 2020). Other orthogonal studies embedded in these surveys are reported in (Guess, Lockett, Lyons, Montgomery, Nyhan and Reifler, 2020) and (Berlinski et al., 2021).

appeared during the 2018 midterms, were published by actual mainstream and false news sources and were balanced within each group in terms of their partisan congeniality. In total, we selected four mainstream news articles that were congenial to Democrats and four that were congenial to Republicans (each split between low- and high-prominence sources), and two pro-Democrat and two pro-Republican false news articles. We define high prominence mainstream sources as those that more than four in ten Americans reported recognizing in recent polling by Pew (Mitchell et al., 2014). False news stories were verified as false by at least one third party fact-checking organization.[8] To the extent possible, we chose stories that were balanced in their face validity. The complete listing of all stories tested is provided in SI Appendix A.

The stories were formatted exactly as they appeared in the Facebook news feed at the time the study was designed. This format replicated the decision environment faced by everyday users, who frequently assess the accuracy of news stories given only the content that appears in social media feeds.[9] Respondents rated twelve stories provided in randomized order during the second wave of each survey.[10]

We then calculate their measured ability to discern mainstream from false news. We do this by taking the difference in the mean perceived accuracy between true and false news headlines (i.e., mean perceived mainstream news accuracy - mean perceived false news accuracy). We use a difference score, rather than perceived accuracy of false news alone, to account for respondents who may tend to rate all news as mostly accurate (i.e., are highly credulous) or all news as mostly inaccurate (i.e., are indiscriminately skeptical). This approach has been frequently used in past studies (e.g., Pennycook and Rand, 2018a; Guess, Lerner, Lyons, Montgomery, Nyhan, Reifler and Sircar, 2020).

SI Appendix Table B1 shows descriptive statistics for the mean perceived accuracy of mainstream and false headlines as well as the difference score in each wave. The results show that, on average, respondents did find mainstream stories to be more credible. For instance, the average rating for mainstream articles in the Oct./Nov. wave was 2.68 while it was 1.90 for false headlines. Although these differences are statistically distinguishable, the *difference* — our measure of discernment — is less than one-point on the four point scale (0.78 for Oct./Nov. and 0.62 for Nov./Dec.). In other words, respondents rated a mainstream headline as less than one point more accurate on our four-point accuracy scale compared to false news headlines. The ranges of values we observe for discernment are -1.5 to 2.88 in Oct./Nov. and -1.38 to 2.75 in Nov./Dec. Although our inferences regarding overconfidence are based on a relatively small number of news headlines ($k = 12$), these headlines appear to be comparable to the large set of political headlines in Pennycook et al. (Pennycook, Binnendyk, Newton and Rand, 2020) ($k = 146$). After re-scaling all outcomes to range from 0–1, the average accuracy rating for our mainstream headlines was .67/.66 across our two surveys and the average rating for false headlines was .48/.50. These mean values are highly

---

[8]Respondents also rated the accuracy of four hyper-partisan news headlines, which are technically factual but present slanted facts in a deceptive manner. We do not include these articles in this analysis due to the inherent ambiguity as to whether they are truthful. These headlines were included as part of a separate study reported in Guess et al. (Guess, Lerner, Lyons, Montgomery, Nyhan, Reifler and Sircar, 2020).

[9]Due to Facebook's native formatting, the visual appearance of the false article previews differed somewhat from those of the mainstream articles — see SI Appendix Figures A1 and A2.

[10]In each survey's first wave, respondents were randomly assigned to evaluate 1 of the 2 stories that fall into each of the 6 categories (e.g., pro-Republican false news, pro-Democrat high-prominence mainstream news, etc.) for a total of six headline evaluations. In the second wave, respondents evaluated all 12 stories using the same approach. We focus only on the wave 2 measures.

similar to Pennycook et al., who found an average rating of .63 for mainstream headlines and .49 for false headlines.

With our discernment measure, we then then order respondents and calculate their percentile. That is, each respondent is scored on a scale ranging from one to one hundred based on their performance where a score of one means that 99% of respondents performed better and a score of 99 means they performed better than 99% of respondents. In the Oct./Nov. survey, the 25th percentile score for discernment was .38, the 50th was .88, the 75th was 1.25, and the 99th was 2.38. Similarly, in Nov./Dec., the 25th percentile discernment score was .13, the 50th was .63 the 75th was 1.00, and the 99th was 2.25.

Accuracy of perceptions of relative ability (overconfidence)

After the headline rating task, we ask two questions in wave 2 of each survey that directly measure differences in perceived ability to detect false news compared to the public.

1. "How do you think you compare to other Americans in your general ability to recognize news that is made up? Please respond using the scale below, where 1 means you're at the very bottom (worse than 99% of people) and 100 means you're at the very top (better than 99% of people),"

2. "How do you think you compare to other Americans in how well you performed in this study at recognizing news that is made up? Please respond using the scale below, where 1 means you're at the very bottom (worse than 99% of people) and 100 means you're at the very top (better than 99% of people)."

For each question, respondents could use a slider to indicate a number between 1 and 100. These measures ("general ability"/"in this study") are highly correlated (r = .73 in both Oct./Nov. 2018 and Nov./Dec. 2018 surveys), so we take their average as our measure of *perceived relative ability*. On the resulting scale, the mean self-assessed relative ability was in the 69th percentile for both surveys (Oct./Nov. M = 69.46, SD = 18.59; Nov./Dec. M = 69.43, SD = 17.8). In both surveys, fewer than 12% of respondents placed themselves below the 50th percentile. The full distributions are shown in SI Appendix Figure B1.

We then combine these variables to compute the overconfidence measure as the difference between people's self-reported ability and their actual performance. The result is a scale that can range from -100 to 100. We show the distribution of overconfidence in SI Appendix B2. (SI Appendix Figure B3 shows the fairly weak relationship between self-rating and actual ability underlying the overconfidence measure.) In both surveys, 73% of respondents were at least somewhat overconfident, with an average overconfidence score of 21.76 in Oct./Nov. and 21.7 in Nov./Dec. (SD = 30.77 to 30.43), meaning the average respondent placed themselves about 22 percentiles higher than their actual score warranted. About 20% of respondents in each survey rated themselves 50 or more percentiles higher than their discernment score warranted.

One potential concern is that our measure of overconfidence may be driven by differences in people's ability to recognize one type of stories rather than how well they can differentiate between them *per se*. SI Appendix Figure B4 therefore disaggregates these components. The figure shows that overconfident respondents perceived mainstream news as less accurate than their counterparts, and, to an even greater extent, perceived false news as more accurate than their counterparts.

Outcomes and behaviors of interest

To answer RQ2–RQ4, we also create measures of visits to false news websites, topical misperceptions, and self-reported engagement intentions (sharing/liking). We describe our measures for each in turn.

News exposure data   News exposure is measured using behavioral data on respondents' web visits collected unobtrusively with their informed consent. Data is available from users' laptop or desktop computers. Web visits are collected anonymously with users' permission through a mix of browser plug-ins, proxies, and VPNs. The provider of this passive metering data is the firm Reality Mine, whose technology underlies the YouGov Pulse panel from which survey respondents were sampled. Our measures of news exposure come from a period immediately following the survey. The lists we used to code each type of media are below:

- Mainstream news visit: One of AOL, ABC News, CBSNews.com, CNN.com, FiveThirtyEight, FoxNews.com, Huffington Post, MSN.com, NBCNews.com, NYTimes.com, Politico, RealClearPolitics, Talking Points Memo, The Weekly Standard, WashingtonPost.com, WSJ.com, or Wikipedia.

- False news visit: Any visit to one of the 673 domains identified (Allcott, Gentzkow and Yu, 2019) as a false news producer as of September 2018 excluding those with print versions (including but not limited to *Express*, the British tabloid) and also domains that were previously classified (Bakshy, Messing and Adamic, 2015) as a source of hard news. In addition, we exclude sites that predominantly feature user-generated content (e.g., online bulletin boards) and political interest groups.

Duplicate visits to webpages were not counted if they were successive (i.e., a page that was reloaded after first opening it). URLs were cleaned of referrer information and other parameters before de-duplication. (For more details, see the processing steps described in Guess (Guess, Nyhan and Reifler, 2020).)

We first created a binary measure of whether respondents made one or more visits to false news sites.[11] Our binary measure of *false news exposure* is coded as 1 if the respondent visited any of the domains in our list (Oct./Nov.: 7%, Nov./Dec.: 6%) and 0 otherwise. We also created a binary measure of *mainstream news exposure* that is coded as 1 if the respondent visited any such domain in our list (Oct./Nov.: 60%, Nov./Dec.: 52%) and 0 otherwise. We use the latter measure to account for the possibility that overconfident individuals may simply be more likely to be exposed to news online.

In addition to false and mainstream news exposure, we also measure the overall ideological slant of respondents' total information diet, which we divide into deciles from most liberal (decile 1) to most conservative (decile 10) using the method presented by Guess (Guess, Forthcoming). We use this measure in our analysis of news exposure to control for the general ideological orientation of respondents' news diets.

---

[11]The distribution was highly skewed; 93–94% of respondents visited zero false news sites and the distribution among non-zero respondents had a long right tail (Oct./Nov. M =.43, SD = 3.24, min=0, max=75; Nov./Dec. M =.21, SD = 1.37, min=0, max=25).

Importantly, not all respondents who were part of our survey chose to provide behavioral data. Thus, our sample sizes using this data decrease, especially in the Nov./Dec. wave in which only 22% of respondents also provided online traffic data (vs. 63% in Oct./Nov.). The decline between surveys reflects the lack of available respondents who (a) participated in YouGov Pulse panel and (b) did not participate in our earlier waves of data collection. The result is that analyses using news exposure data have less power (we also consider pooled analyses across surveys for this reason).

Topical misperceptions, engagement, and congeniality   In the Oct./Nov. survey, we included a battery of questions asking respondents about their beliefs in specific claims related to the confirmation hearings for Justice Brett Kavanaugh, which occurred shortly before the survey was fielded.[12] Respondents were shown two true and two false statements that they rated on a four-point accuracy scale ranging from "Not at all accurate" to "Very accurate."

These statements were balanced in terms of partisan orientation so one true and one false statement was congenial to Democrats and one true and one false statement was congenial to Republicans. Both the true and false statements were highly visible on social media during the hearings.[13] We measured potential engagement (liking/sharing) with false news stories during the headline rating task. For each headline, respondents were asked to self-report their intention to "like" or "share" each article (1 = not at all likely, 4 = very likely). This question was asked only of respondents who report using Facebook.[14] It should be noted that perceived accuracy questions appeared immediately before our sharing intent questions in the survey, which may prime accuracy concerns among respondents and thereby alter self-reported sharing behavior (Pennycook, Epstein, Mosleh, Arechar, Eckles and Rand, 2020).

For both misperceptions and engagement, we analyze the data in two ways. First, we create a *difference score*. For the topical misperceptions, for instance, we subtracted the perceived accuracy of false statements from the perceived accuracy of true statements to create a measure of discernment. We calculated mean responses of intentions to like/share mainstream stories, false stories, and the difference using an identical procedure. Descriptive statistics for these measures are shown in SI Appendix Table B1.

Finally, we also examine our results at the headline or statement level using only false statements/headlines so that we can test whether the relationship between overconfidence and beliefs or behavior varies by partisan congeniality. Congeniality is coded at the headline or statement level for partisans to indicate that a story or statement is consistent with the respondents' partisan leanings (e.g., a Democrat evaluating a story that is favorable to a Democrat). To determine the partisanship of respondents in the U.S. survey, we used the standard two-question party identification battery (which includes leaners) to classify respondents as Democrats or Republicans.

---

[12]The confirmation hearings where Kavanaugh and Christine Blasey Ford testified took place in late September 2018. The final senate vote took place on October 6.

[13]This battery was included in both waves of the Oct./Nov. survey. We focus only on the wave 2 results as the first wave preceded our collection of the overconfidence measure, but SI Appendix D shows that our results replicate fully when using the wave 1 topical misperception battery.

[14]We only observe self-reported behavioral intentions. However, self-reported sharing intention for political news articles has been shown to correlate with aggregate observed sharing behavior on Twitter at $r = .44$ (Mosleh, Pennycook and Rand, 2020).

Additional covariates   Our statistical models include a series of standard covariates including dichotomous indicators of Democrat and Republican party affiliation (including leaners), college education, gender, nonwhite racial background, and dichotomous indicators of membership in age groups (30–44, 45–59, and 60+; 18–29 is the omitted category). Complete descriptions of all survey items and measures are included in SI Appendix A.

Our October/November 2018 respondents are 57% female, 80% white, median age 55, 37% hold a four-year college degree or higher, 49% identify as Democrats (including leaners), and 34% identify as Republicans (including leaners). Our November/December 2018 respondents are 55% female, 68% white, median age 50, 32% hold a four-year college degree or higher, 46% identify as Democrats (including leaners), and 36% identify as Republicans (including leaners).

Data availability

Data files and scripts necessary to replicate the results in this article will be made available at the following Open Science Framework repository: `https://osf.io/xygwt/`.

References

Allcott, Hunt, Matthew Gentzkow and Chuan Yu. 2019. "Trends in the diffusion of misinformation on social media." *Research & Politics* 6(2):2053168019848554.

Anderson, Cameron, Sebastien Brion, Don A Moore and Jessica A Kennedy. 2012. "A status-enhancement account of overconfidence." *Journal of personality and social psychology* 103(4):718.

Anson, Ian G. 2018. "Partisanship, political knowledge, and the dunning-kruger effect." *Political Psychology* 39(5):1173–1192.

Bago, Bence, David G Rand and Gordon Pennycook. 2020. "Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines." *Journal of experimental psychology: general* .

Bakshy, Eytan, Solomon Messing and Lada A. Adamic. 2015. "Exposure to ideologically diverse news and opinion on Facebook." *Science* 348(6239):1130–1132.

Barthel, Michael, Amy Mitchell and Jesse Holcomb. 2016. "Many Americans believe fake news is sowing confusion." *Pew Research Center* 15:12.

Belmi, Peter, Margaret A Neale, David Reiff and Rosemary Ulfe. 2020. "The social advantage of miscalibrated individuals: The relationship between social class and overconfidence and its implications for class-based inequality." *Journal of personality and social psychology* 118(2):254.

Berlinski, N, Margaret Doyle, A Guess, G Levy, B Lyons, J Montgomery, B Nyhan and J Reifler. 2021. "The Effects of Unsubstantiated Claims of Voter Fraud on Confidence in Elections." *Journal of Experimental Political Science* .

Brashier, Nadia M and Elizabeth J Marsh. 2020. "Judging truth." *Annual review of psychology* 71.

Bregu, Klajdi. 2020. "Overconfidence and (Over) Trading: The Effect of Feedback on Trading Behavior." *Journal of Behavioral and Experimental Economics* 88:101598.

Burson, Katherine A, Richard P Larrick and Joshua Klayman. 2006. "Skilled or unskilled, but still unaware of it: how perceptions of difficulty drive miscalibration in relative comparisons." *Journal of personality and social psychology* 90(1):60.

Cheng, Joey T, Cameron Anderson, Elizabeth R Tenney, Sebastien Brion, Don A Moore and Jennifer M Logg. 2020. "The social transmission of overconfidence." *Journal of Experimental Psychology: General* .

Davison, W Phillips. 1983. "The third-person effect in communication." *Public opinion quarterly* 47(1):1–15.

Douglas, Karen M and Robbie M Sutton. 2004. "Right about others, wrong about ourselves? Actual and perceived self-other differences in resistance to persuasion." *British Journal of Social Psychology* 43(4):585–603.

Duflo, Esther, Abhijit Banerjee, Amy Finkelstein, Lawrence F Katz, Benjamin A Olken and Anja Sautmann. 2020. In praise of moderation: Suggestions for the scope and use of pre-analysis plans for rcts in economics. Technical report National Bureau of Economic Research.

Dunning, David. 2011. "The Dunning–Kruger effect: On being ignorant of one's own ignorance." *Advances in experimental social psychology* 44:247–296.

Ehrlinger, Joyce, Kerri Johnson, Matthew Banner, David Dunning and Justin Kruger. 2008. "Why the unskilled are unaware: Further explorations of (absent) self-insight among the incompetent." *Organizational behavior and human decision processes* 105(1):98–121.

Feld, Jan, Jan Sauermann and Andries De Grip. 2017. "Estimating the relationship between skill and overconfidence." *Journal of behavioral and experimental economics* 68:18–24.

Ferraro, Paul J. 2010. "Know thyself: Competence and self-awareness." *Atlantic Economic Journal* 38(2):183–196.

Flynn, DJ, Brendan Nyhan and Jason Reifler. 2017. "The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics." *Political Psychology* 38:127–150.

Gignac, Gilles E and Marcin Zajenkowski. 2020. "The Dunning-Kruger effect is (mostly) a statistical artefact: Valid approaches to testing the hypothesis with individual differences data." *Intelligence* 80:101449.

Graham, Matthew H. 2020. "Self-awareness of political knowledge." *Political Behavior* 42(1):305–326.

Grinberg, Nir, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson and David Lazer. 2019. "Fake news on Twitter during the 2016 US presidential election." *Science* 363(6425):374–378.

Guess, Andrew M. Forthcoming. "(Almost) Everything in Moderation: New Evidence on Americans' Online Media Diets." *American Journal of Political Science* .

Guess, Andrew M, Brendan Nyhan and Jason Reifler. 2020. "Exposure to untrustworthy websites in the 2016 US election." *Nature human behaviour* 4(5):472–480.

Guess, Andrew M, Dominique Lockett, Benjamin Lyons, Jacob M Montgomery, Brendan Nyhan and Jason Reifler. 2020. ""Fake news" may have limited effects beyond increasing beliefs in false claims." *Harvard Kennedy School Misinformation Review* 1(1).

Guess, Andrew M, Michael Lerner, Benjamin Lyons, Jacob M Montgomery, Brendan Nyhan, Jason Reifler and Neelanjan Sircar. 2020. "A digital media literacy intervention increases discernment between mainstream and false news in the United States and India." *Proceedings of the National Academy of Sciences* 117(27):15536–15545.

Hansen, Eric M, Kremena Yakimova, Marie Wallin and Linda Thomsen. 2010. Can thinking you're skeptical make you more gullible? The illusion of invulnerability and resistance to manipulation. In *The individual and the group: Future challenges*, ed. Christian Jacobsson and Max Rapp Ricciardi. Proceedings from the 7th GRASP conference, University of Gothenburg, May 2010.

Kovacs, Roxanne J, Mylene Lagarde and John Cairns. 2020. "Overconfident health workers provide lower quality healthcare." *Journal of Economic Psychology* 76:102213.

Krueger, Joachim and Ross A Mueller. 2002. "Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance." *Journal of personality and social psychology* 82(2):180.

Kruger, Justin and David Dunning. 1999. "Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments." *Journal of personality and social psychology* 77(6):1121.

Kruger, Justin and David Dunning. 2002. "Unskilled and unaware–but why? A reply to Krueger and Mueller (2002)." *Journal of Personality and Social Psychology* 82(2):189–192.

Lambert, Jérôme, Véronique Bessière and Gilles N'Goala. 2012. "Does expertise influence the impact of overconfidence on judgment, valuation and investment decision?" *Journal of Economic Psychology* 33(6):1115–1128.

Li, Jianing and Michael W Wagner. 2020. "The Value of Not Knowing: Partisan Cue-Taking and Belief Updating of the Uninformed, the Ambiguous, and the Misinformed." *Journal of Communication* .

Martel, Cameron, Gordon Pennycook and David G Rand. 2020. "Reliance on emotion promotes belief in fake news." *Cognitive research: principles and implications* 5(1):1–20.

McIntosh, Robert D, Elizabeth A Fowler, Tianjiao Lyu and Sergio Della Sala. 2019. "Wise up: Clarifying the role of metacognition in the Dunning-Kruger effect." *Journal of Experimental Psychology: General* .

Miller, Jane E, Paul D Windschitl, Teresa A Treat and Aaron M Scherer. 2019. "Unhealthy and unaware? Misjudging social comparative standing for health-relevant behavior." *Journal of Experimental Social Psychology* 85:103873.

Mitchell, Amy, Jeffrey Gottfried, Jocelyn Kiley and Katerina Eva Matsa. 2014. "Political polarization & media habits." Pew Research Center, October 21, 2014. Downloaded March 21, 2019 from `https://www.pewresearch.org/wp-content/uploads/sites/8/2014/10/Political-Polarization-and-Media-Habits-FINAL-REPORT-7-27-15.pdf`.

Mondak, Jeffery J and Mary R Anderson. 2004. "The knowledge gap: A reexamination of gender-based differences in political knowledge." *The Journal of Politics* 66(2):492–512.

Mosleh, Mohsen, Gordon Pennycook and David G Rand. 2020. "Self-reported willingness to share political news articles in online surveys correlates with actual sharing on Twitter." *Plos one* 15(2):e0228882.

Motta, Matthew, Timothy Callaghan and Steven Sylvester. 2018. "Knowing less but presuming more: Dunning-Kruger effects and the endorsement of anti-vaccine policy attitudes." *Social Science & Medicine* 211:274–281.

Niederle, Muriel and Lise Vesterlund. 2007. "Do women shy away from competition? Do men compete too much?" *The quarterly journal of economics* 122(3):1067–1101.

Ortoleva, Pietro and Erik Snowberg. 2015*a*. "Are conservatives overconfident?" *European Journal of Political Economy* 40:333–344.

Ortoleva, Pietro and Erik Snowberg. 2015*b*. "Overconfidence in political behavior." *American Economic Review* 105(2):504–35.

Parker, Andrew M and Eric R Stone. 2014. "Identifying the effects of unjustified confidence versus overconfidence: Lessons learned from two analytic methods." *Journal of behavioral decision making* 27(2):134–145.

Pasek, Josh, Gaurav Sood and Jon A Krosnick. 2015. "Misinformed about the affordable care act? Leveraging certainty to assess the prevalence of misperceptions." *Journal of Communication* 65(4):660–673.

Pazicni, Samuel and Christopher F Bauer. 2014. "Characterizing illusions of competence in introductory chemistry students." *Chemistry Education Research and Practice* 15(1):24–34.

Pennycook, Gordon and David G. Rand. 2018*a*. "Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning." *Cognition* .

Pennycook, Gordon and David G Rand. 2018*b*. "Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking." *Journal of personality* .

Pennycook, Gordon, Jabin Binnendyk, Christie Newton and David Rand. 2020. "A practical guide to doing behavioural research on fake news and misinformation." PsyArXiv [Preprint] (2020). https://psyarxiv.com/g69ha (accessed 21 January 2021).

Pennycook, Gordon, Robert M Ross, Derek J Koehler and Jonathan A Fugelsang. 2017. "Dunning–Kruger effects in reasoning: Theoretical implications of the failure to recognize incompetence." *Psychonomic bulletin & review* 24(6):1774–1784.

Pennycook, Gordon, Ziv Epstein, Mohsen Mosleh, Antonio A Arechar, Dean Eckles and David Rand. 2020. "Shifting attention to accuracy can reduce misinformation online." PsyArXiv [Preprint] (2020). https://psyarxiv.com/3n9u8/ (accessed 21 January 2021).

Perloff, Richard M. 1989. "Ego-involvement and the third person effect of televised news coverage." *Communication research* 16(2):236–262.

Prims, Julia P and Don A Moore. 2017. "Overconfidence over the lifespan." *Judgment and decision making* 12(1):29.

Roozenbeek, Jon and Sander van der Linden. 2019. "Fake news game confers psychological resistance against online misinformation." *Palgrave Communications* 5(1):1–10.

Ross, Lee, David Greene and Pamela House. 1977. "The 'false consensus effect': An egocentric bias in social perception and attribution processes." *Journal of experimental social psychology* 13(3):279–301.

Salovich, Nikita A and David N Rapp. 2020. "Misinformed and unaware? Metacognition and the influence of inaccurate information." *Journal of experimental psychology: learning, memory, and cognition* .

Schlösser, Thomas, David Dunning, Kerri L Johnson and Justin Kruger. 2013. "How unaware are the unskilled? Empirical tests of the "signal extraction" counterexplanation for the Dunning–Kruger effect in self-evaluation of performance." *Journal of Economic Psychology* 39:85–100.

Sheffer, Lior and Peter Loewen. 2019. "Electoral confidence, overconfidence, and risky behavior: Evidence from a study with elected politicians." *Political Behavior* 41(1):31–51.

Sheldon, Oliver J, David Dunning and Daniel R Ames. 2014. "Emotionally unskilled, unaware, and uninterested in learning more: Reactions to feedback about deficits in emotional intelligence." *Journal of Applied Psychology* 99(1):125.

Slater, Michael D. 2007. "Reinforcing spirals: The mutual influence of media selectivity and media effects and their impact on individual behavior and social identity." *Communication theory* 17(3):281–303.

Stankov, Lazar and Jihyun Lee. 2014. "Overconfidence across world regions." *Journal of Cross-Cultural Psychology* 45(5):821–837.

Sun, Ye, Lijiang Shen and Zhongdang Pan. 2008. "On the behavioral component of the third-person effect." *Communication Research* 35(2):257–278.

Thompson, Valerie A, Jamie A Prowse Turner and Gordon Pennycook. 2011. "Intuition, reason, and metacognition." *Cognitive psychology* 63(3):107–140.

Vegetti, Federico and Moreno Mancosu. 2020. "The Impact of Political Sophistication and Motivated Reasoning on Misinformation." *Political Communication*.
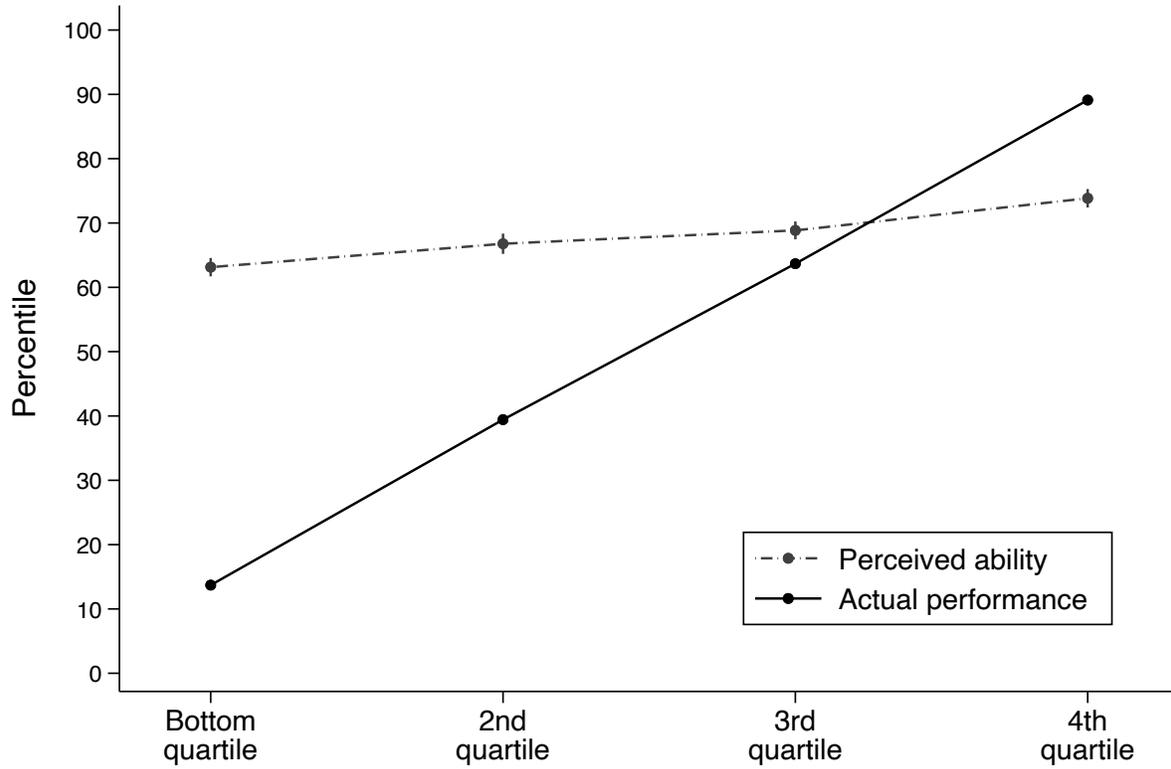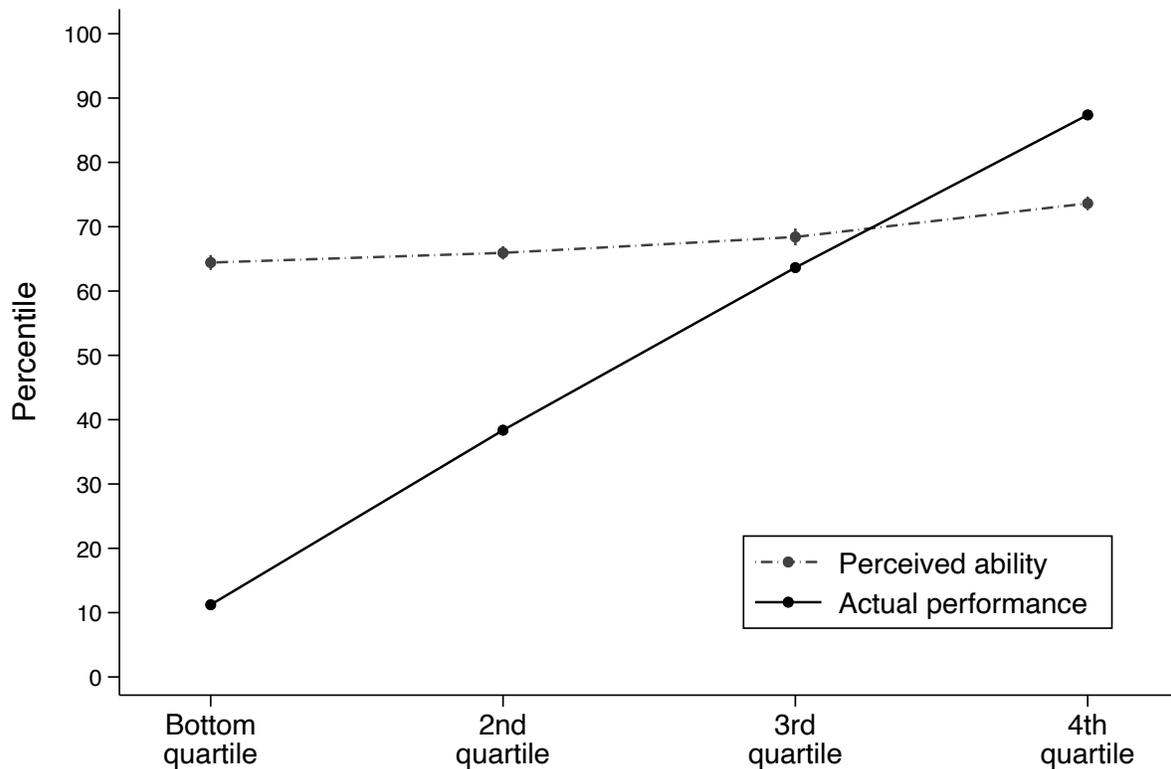
Figure 1: Perceived false news detection ability for respondents grouped by actual performance. Notes: Gaps depict miscalibration between actual and self-assessed percentile of performance for quartile groups based on actual performance with 95% CIs (note: CIs are smaller than the markers for actual performance and thus not visible). Oct./Nov. N = 2,855, Nov./Dec. N = 4,150.
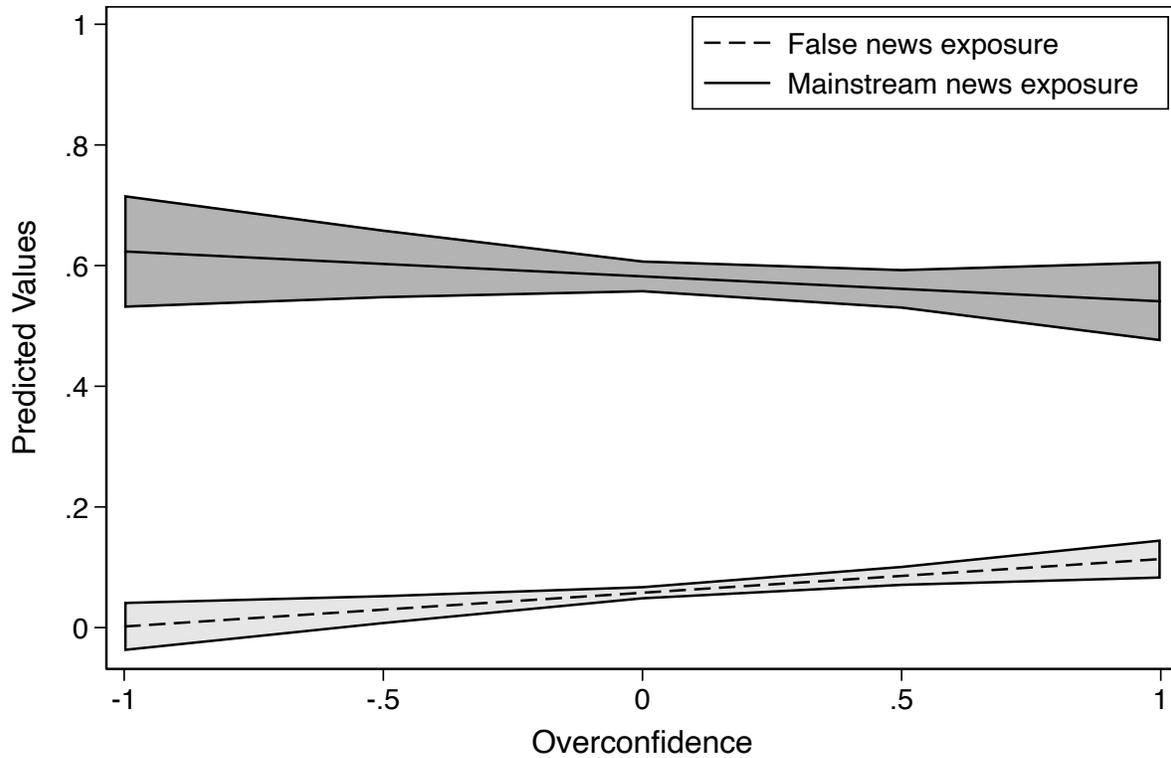
Figure 2: Overconfidence and news exposure. Notes: Predictive margins with 95% confidence intervals, based on full model, all other variables held constant. The overconfidence measure subtracts respondents' actual percentile from their self-rated percentile and is re-scaled to range from -1 to 1. False news exposure is coded as 1 if the respondent visited any such domain and 0 otherwise. Mainstream news exposure is coded as 1 if the respondent visited any such domain and 0 otherwise. Data come from the pooled model (N=2,547) which pools data from Oct./Nov. (N=1780) and Nov./Dec. (N=767) surveys.