

# Influences of Source Bias That Differ From Source Untrustworthiness: When Flip-Flopping Is More and Less Surprising

Laura E. Wallace, Duane T. Wegener, and Richard E. Petty  
The Ohio State University

Discussions of the difference between biased and fake news were prevalent after the 2016 United States Presidential election. However, within social psychology, and especially the psychology of persuasion, perceptions of source bias have been largely overlooked or conflated with untrustworthiness. In the current work, we sought to demonstrate that bias and untrustworthiness can have differing effects. One such situation is when persuasive sources originally take one position but switch to a different position (flip-flopping). We find that people expect biased versus objective sources to consistently maintain their position. Conversely, people do not have these expectations for untrustworthy versus trustworthy sources. When sources unexpectedly switch positions, people can infer that they must have switched because of strong evidence in support of the new position. As a result, taking an unexpected position can lead a source to be more persuasive. This package includes a final study with a preregistered analysis plan that uses latent variable modeling, as well as an integrative data analysis across all data we have to test these hypotheses. Ultimately, this work suggests that bias and untrustworthiness can have differing indirect influences on persuasion when sources switch positions, highlighting the need to conceptually separate bias and untrustworthiness and examine their individual effects. These persuasive effects function as an illustrative example of differing influences of bias and untrustworthiness, but we expect this distinction to have theoretical implications across domains of social psychology and practical applications for media producers and consumers.

**Keywords:** bias, trustworthiness, persuasion, expectations, flip-flopping

**Supplemental materials:** <http://dx.doi.org/10.1037/pspa0000181.supp>

From early in the 2016 United States Presidential election, Republican candidate Donald Trump used the term “fake news” to discredit the mainstream media. The 2016 election also contained a lot of fake news in that there were several news outlets publishing false election-related stories. These events led to a flurry of discussion about the distinction between “biased” and “fake” news and whether this difference matters. Given the long history of person perception and persuasion research in social psychology, it

seems reasonable to assume that at least one of these literatures would have a clear answer to the question of the difference between perceiving sources as biased versus fake (untrustworthy). However, until recently, these concepts have typically been conflated with one another. Despite this confusion, recent research has separated source bias (skewed perception) from source untrustworthiness (dishonesty), and demonstrated independent, but similar consequences of each perception. For example, each can have independent negative effects on persuasion and perceptions of credibility (Wallace, Wegener, & Petty, 2019). However, it remains unclear whether bias and untrustworthiness can have *differing* consequences from one another, a question we examine in the current article.

---

 Laura E. Wallace, Duane T. Wegener, and Richard E. Petty, Department of Psychology, The Ohio State University.

This work was supported by the National Science Foundation Graduate Research Fellowship Program (DGE-1343012). Studies 1 and 2 were presented at the 2018 Attitudes Preconference at the annual meeting of the Society for Personality and Social Psychology, Atlanta, GA and at the 2018 Ohio State Symposium on Social Psychology in Columbus, OH. An integrative data analysis of Studies 1–6 was presented at the 2019 Attitudes Preconference at the annual meeting of the Society for Personality and Social Psychology. Comments by members of the 2013–2018 Groups for Attitudes and Persuasion at The Ohio State University are appreciated.

Link to preregistration on OSF: [https://osf.io/m38vk/?view\\_only=048d25c77c3946a1915797885637c9a4](https://osf.io/m38vk/?view_only=048d25c77c3946a1915797885637c9a4).

Correspondence concerning this article should be addressed to Laura E. Wallace, Department of Psychology, The Ohio State University, 1835 Neil Avenue, Columbus, OH 43210. E-mail: [wallace.748@osu.edu](mailto:wallace.748@osu.edu)

## Historical Confusion About Perceptions of Bias

Despite researchers across domains of social psychology attempting to identify and distinguish certain person perceptions expected to be particularly influential, research examining perceptions of bias has generally been sparse and remained isolated from other work on person perception more generally (Kennedy & Pronin, 2008; Pronin, Gilovich, & Ross, 2004; Ross & Ward, 1996). For example, work on the stereotype content model (Cuddy, Fiske, & Glick, 2008) demonstrated that people tend to stereotype others along the dimensions of warmth and competence. Similarly, persuasion researchers (Hovland, Janis, & Kelley, 1953) conceptualized source credibility as composed of trustworthiness and

expertise. To the extent that persuasion researchers have studied perceptions of bias, historically they have mostly done so by examining effects of source vested interest (when sources have something to gain through their advocacy), demonstrating that sources without a vested interest were more persuasive and perceived as more credible (e.g., Hovland & Mandell, 1952; Walster, Aronson, & Abrahams, 1966). Researchers referred to these manipulations of vested interest interchangeably as manipulating source bias or untrustworthiness. Consistent with this, these manipulations likely would have affected perceptions of both bias and untrustworthiness (for evidence of this, see Wallace, 2019), making it difficult to draw conclusions about the effects of bias on persuasion separate from effects of untrustworthiness.

### Distinguishing Bias From Untrustworthiness

To more clearly distinguish between bias and untrustworthiness, recent research has experimentally demonstrated that participants conceptualize bias as a motivation to hold a particular position and untrustworthiness as willingness to be dishonest (Wallace et al., 2019). Further, people can naturally distinguish between these perceptions, at times inferring bias without inferring untrustworthiness. Additionally, perceiving a source as biased can have an independent negative effect on perceptions of source credibility and persuasion beyond any effects of source trustworthiness and expertise (Wallace et al., 2019). Consistent with some findings from Eagly, Wood, and Chaiken (1978), Wallace (2015) demonstrated that bias loads separately from untrustworthiness in exploratory factor analyses. Thus, this recent work has provided evidence to separate bias from untrustworthiness and demonstrated that bias can have independent influences on perceptions of source credibility and on persuasion.

Consistent with this recent work, one early exception to the general tendency to either overlook source bias or conflate it with untrustworthiness was research by Eagly et al. (1978). Their studies examined what they referred to as knowledge bias (a slant in knowledge because of asymmetrical experience with an issue) and reporting bias (that they linked directly to untrustworthiness and implemented by suggesting that the person was, in one condition, saying what they truly believe and, in another condition, presenting a position they did not believe—that is, the source was dishonest). The primary finding of the article was that violating expectations, regardless of whether those expectations were based on a potential knowledge bias or reporting bias, generally led to increased persuasion. The explanation for this was attributional. When a source takes an expected position, people can attribute it to the source's previous knowledge or to the audience's desires, depending on how the expectation was established. However, when the source takes the opposite position, people may attribute the source's stance to strong evidence supporting this unexpected position.

### Hypothesized Differing Consequences of Source Bias and Untrustworthiness

Although evidence has accumulated that bias and untrustworthiness are conceptually distinct and can have separate effects, the research has, up to this point, solely demonstrated that bias and untrustworthiness can have the *same* consequences as one another

(e.g., biased and untrustworthy sources each decreasing persuasion). We hypothesized, however, that there may be times when bias has directionally distinct effects from untrustworthiness. Specifically, we predicted that perceiving a source as biased should lead participants to expect the source to maintain the same position on topics related to that bias over time. Conversely, source untrustworthiness should have a lesser or even opposite effect on these expectations. For example, imagine that people identify a news source as having an ideological bias.<sup>1</sup> People would then expect the news source to consistently print information in support of their ideological position.

Yet, simply knowing that a source is untrustworthy should not necessarily lead people to develop expectations about the position a source will take. If anything, an untrustworthy source might be particularly unpredictable. Supermarket tabloids, such as *Weekly World News* and the *National Enquirer*, are famous for making outrageous claims such as Stevie Wonder regaining sight or the existence of Bat Boy (half human, half bat). These outlets represent excellent examples of untrustworthy sources. If one simply knows that a tabloid reports false news, it would be challenging to predict in advance whether that information would support one position or the other. Relatedly, it is not clear that one would expect any consistency in positions taken over time, as untrustworthy sources are not even bound by the truth. Conversely, trustworthy sources are bound by the truth. Because of this, we thought it possible that untrustworthiness would have a negative effect on expectations that the source would switch positions. However, it may also be challenging to predict the position a trustworthy source would take because they will also switch positions when the information justifies such changes. Thus, we hypothesized that we would observe a negative or null effect of source untrustworthiness on expectations of consistency.

Understanding the influence of source characteristics on expectations about the position the source will take may be important because violation of these expectations could increase persuasion (Eagly et al., 1978). As such, the goal of the current work was to demonstrate that bias and untrustworthiness can have different (potentially opposite) effects on expectations about the position the source would take, which could result in opposing downstream influences on persuasion.

Although the current work clearly builds on research conducted by Eagly et al. (1978), the goals and methods of the work differ in several important ways. Most importantly, in the current work, we were interested in how source characteristics (our independent variables) might influence inferences about the position the source would take next. Thus, in our studies, participants learned about the source's bias or untrustworthiness before the message, and we used these perceptions to establish different expectations about what a shift in position might mean. Conversely, the Eagly et al. (1978) research was aimed at examining inferences of knowledge bias or reporting bias (untrustworthiness) from the positions the sources took (i.e., bias and untrustworthiness were the dependent

<sup>1</sup> As noted, a source is biased when the source is motivated to hold a particular position. This position could stem from one's politics, religion, self or family interest, business ties, and many other factors. A source could also be biased just on one issue (e.g., one's grandchildren), on some set of thematically connected issues (e.g., ideological issues), or on every issue (e.g., because of one's self interest).

variables). This meant that participants did not have perceptions of the source as (un)biased or (un)trustworthy before learning the position of the target message. Thus, we do not believe that Eagly et al. (1978) tested the current hypotheses regarding potential differences between source bias and source untrustworthiness, nor were they designed to do so. Further, we extend the Eagly et al. findings by explicitly testing the hypothesis that unexpected position switching can lead to inferences that the source has particularly strong reasons to support the new position.

### Overview of Studies

We report seven studies that test our hypotheses. To efficiently present results, we present combined analyses across studies that used a similar design. However, parallel analyses, along with means, standard deviations, and a covariance matrix for each study are available in the online supplemental materials. Studies 1, 2, and 3 crossed bias and untrustworthiness manipulations to examine experimental effects of bias and untrustworthiness on surprise at position switching, as well as downstream consequences of this surprise for inferences about information quality, and ultimately, persuasion. Studies 4, 5, and 6 were conducted to test the same hypotheses, but they addressed several concerns with the bias manipulation used in the initial studies. We also conducted a highly powered preregistered replication (Study 7) to address any potential concerns about power or researcher degrees of freedom, though, as we explain, any changes across studies were designed to specifically address issues raised in previous studies, so there was no selectivity in measure choice or study reporting. Finally, we present an integrative data analysis of each set of studies and of all the data together to provide an estimate of the effects across all the data available to us. To be clear, the studies reported in this article comprise the only complete studies (not including pretesting) run to test our mediational model, so the integrative model is unaffected by selective reporting of results. Additionally, two studies solely testing the effect of source bias on surprise (with no test of downstream consequences) are available in the online supplemental materials. An integrative data analysis that includes all studies reported in this article plus the two studies without downstream consequences, two pretesting data collections (total  $N = 230$ ), and one pilot study ( $N = 114$ ) is also reported in the online supplemental materials and supports the same conclusions as the analyses in the rest of the article.

### Studies 1, 2, and 3

In Studies 1, 2, and 3, we began by examining whether bias and untrustworthiness might have differing effects on how surprising it would be for a source to switch positions. Specifically, we predicted that perceiving a source as biased versus objective would lead participants to infer that the source would be more likely to maintain the same position on topics relevant to that bias. Conversely, untrustworthiness should have null or possibly opposite effects on these expectations of consistency. Mirroring many real-life situations in which politicians or media outlets are referred to as “biased” or “dishonest/untrustworthy,” in these initial studies, we used direct manipulations characterizing a source as biased versus objective and untrustworthy versus trustworthy.

Additionally, we designed these studies so that we could test whether surprise at position switching would ultimately lead to

increased persuasion. Though not tested previously, expectancy-violation-based increases in persuasion presumably occurred because when the source took an expected position, participants attributed the position to the factor that created the expectation (e.g., the source’s bias; Eagly et al., 1978). Conversely, when the source took an unexpected position, participants attributed the position to the source having learned compelling information supporting the new position. Therefore, when a biased source provides arguments for a new position that opposes their initial position, this should be more unexpected (surprising) than when an objective source switches positions. Thus, the information supporting the new position should then be viewed as *more* persuasive when it is more unexpected (surprising). If the dimension of untrustworthiness does not affect surprise at position switching, then it should also not influence perceptions of the persuasiveness of information supporting the new position.

Because Studies 1, 2, and 3 were almost identical, we present combined results for the three studies, but we describe the relatively small study design differences here. Study 2 was identical to Study 1 except for the inclusion of a weak argument condition. Although we predicted that surprise would influence persuasion through the inference of information quality mechanism originally described by Eagly et al. (1978), we thought it was possible that surprise at the position switch could increase processing of the new message as surprise has led to increased information processing in past research (e.g., Baker & Petty, 1994). If so, then surprise could have a positive effect on persuasion with strong arguments, but a negative effect on persuasion with weak arguments. If the inference process outlined by Eagly et al. (1978) drives perceptions of information quality, however, then surprise could have a positive effect on persuasion in both argument quality conditions. When we compared structural equation models (SEMs) in which the coefficients of the surprise to information quality paths were constrained to be equal across argument quality conditions or were left free to vary, a chi-square difference test provided little evidence of any differences between the argument quality conditions,  $\chi^2(1) = .24$ ,  $p = .62$  (analyses in the online supplemental materials), providing evidence against an amount of processing explanation. Because surprise had the same effect in each condition, we collapsed across the two argument quality conditions when combining with the other studies.

Study 3 was identical to Study 1 except for the inclusion of a credibility measure in Study 3. We included that measure because we thought that even though perceived bias might have a positive effect on persuasion through surprise at position switching, it might have a negative effect on persuasion because of reduced perceptions of credibility for biased sources (Wallace et al., 2019). This possibility suggests that source bias could have opposing influences on persuasion through the surprise and credibility pathways. Although we describe the study methods below, the exact materials for every study are also available in the online supplemental materials. Every study in this article was approved by the Ohio State University Institutional Review Board.

### Method

**Participants.** There were 546 (147 in Study 1, 118 in Study 2, and 281 in Study 3) Amazon Mechanical Turk (MTurk) workers who were recruited to participate in the studies. Our *a priori*

exclusion criterion for the original (first six) studies in this line of work was to exclude those who self-reported not taking the study seriously (a 5 on a 5-point scale). For these three studies, this resulted in excluding two participants.

**Design and procedure.** The studies used a 2 (Bias: high vs. low)  $\times$  2 (Trustworthiness: high vs. low) design so we could examine independent effects of bias and untrustworthiness on expectations about future positions the source will take. After consenting to participate in the study, participants read about a university service program proposed in Canada through which students could receive reduced tuition. They read that a website, APL News, had opposed the university service program. Participants were also told that the university service program had become a heated political issue in Canada, with one side supporting it and the other opposing it, thereby linking the issue to any potential ideological bias the source might have. Then they read a description of APL News from “Media Reports,” which they were told was an independent objective agency concerned with ensuring that people got their news from reputable sources. They read that Media Reports rates news organizations along two dimensions: how objective versus biased they are and how truthful versus untruthful they are. Then they read how Media Reports had rated APL News on each dimension (allowing us to cross bias and untrustworthiness). Participants completed manipulation checks on source bias and untrustworthiness. Additionally, in Study 3, we measured how much participants viewed the source as credible. Then participants reported the position they expected the source to take and how surprised they would be if the source switched positions. They then received a message from APL News favoring the university service program, representing a position switch (message in the online supplemental materials). After reading the message, participants reported how much they thought the source had good reasons for supporting the university service program and their own attitudes toward the program. Finally, they were debriefed and thanked for their participation.

#### Independent variables.

**Source bias.** Participants read that Media Reports rates news agencies along the dimension of how “Objective versus Biased: How objective versus ideologically biased they are.” They read that news organizations were given a rating from  $-5 = \textit{very biased}$  to  $5 = \textit{very objective}$ . Then, when participants received information about APL News, APL News was rated as  $-4 = \textit{very biased}$  in the high-bias condition and  $4 = \textit{very objective}$  in the low-bias condition.

**Source trustworthiness.** Participants also read that Media Reports rates news agencies along the dimension of how “Truthful versus Untruthful: How honest (committed to reporting facts) versus dishonest (willing to report false information)” they are. They read that news organizations were given a rating from  $-5 = \textit{very untruthful}$  to  $5 = \textit{very truthful}$ . Then, when participants received information about APL News, APL News was rated as  $-4 = \textit{very untruthful}$  in the low-trustworthiness condition and  $4 = \textit{very truthful}$  in the high-trustworthiness condition.

Beyond the rating scales, participants received a description of the source. In the high-bias, high-trustworthiness condition, participants read, “APL News reporters are highly committed to reporting the truth, but they can only see the world through their ideological lens.” In the low-bias, low-trustworthiness condition, they read, “APL News has no identifiable political leaning. Some-

times they take liberal positions; sometimes they take conservative positions. On previous occasions, APL News has reported information that they knew was false.” In the low-bias, high-trustworthiness condition they read, “APL News has no identifiable political leaning. Sometimes they take liberal positions; sometimes they take conservative positions. APL News has reporters who are committed to reporting the truth.” Finally, in the high-bias, low-trustworthiness condition, they read, “APL News has reporters who can only view the world through their ideological lens. On previous occasions, APL News has reported information that they knew was false.”

#### Manipulation checks.

**Perceptions of bias.** Participants were asked, “How much do you think APL News is ideologically driven?” (1 = *not at all ideologically driven*, 9 = *very ideologically driven*) and “How much do you think the position APL News took on the university service program was motivated by their ideology?” (1 = *not at all motivated by their ideology*, 9 = *very much motivated by their ideology*),  $r = .88$ .

**Perceptions of trustworthiness.** Participants were asked, “How much do you see APL News as truthful?” (1 = *not at all truthful*, 9 = *very truthful*) and “How much do you perceive APL News as honest?” (1 = *very dishonest*, 9 = *very honest*),  $r = .90$ .

#### Dependent measures.

**Perceptions of credibility.** In Study 3, credibility was measured by asking participants “How much do you see APL News as a credible source?” and “To what extent do you view APL News as a credible source of information on the university service program?” on 9-point scales (1 = *very noncredible*, 9 = *very credible*),  $r = .94$ . Credibility was not assessed in Studies 1 and 2.

**Surprise at position switch.** Participant surprise was composed through a combination of the strength of expectations that the source’s position would remain consistent over time along with direct reports of how surprising a change of position would be,  $\alpha = .82$ . Participants were asked two items intended to measure expectations about the position the source would take, (i.e., “If APL News were to publish an article about the university service program, to what extent do you think it would be consistent or inconsistent with their previous opposition?”; 1 = *I would expect it to be inconsistent*, 5 = *I would not know what to expect*, 9 = *I would expect it to be consistent*). Then after each of these expectations items, they were asked how certain they were in the expectation (i.e., “How certain are you in predicting the stance APL News would take toward the university service program in a future article?”; 1 = *not at all certain*, 9 = *very certain*). After responding to these expectation items, participants learned that APL News was planning to publish an article titled, “Report suggests university service plan would yield many benefits.” Then participants responded to two questions to assess their surprise at the position switch (i.e., “To what extent are you surprised that APL News is publishing an article supporting the university service program?”; 1 = *not at all surprised*, 9 = *very surprised*).

As noted below, we take a latent variable approach to analyzing our results. In each of our analyses, we allowed the residual variances of each of the pairs of items making up the surprise index (expectations, expectation certainty, and surprise) to correlate with one another. We did this because the items in these pairs may be more similar to one another than they are to the other items included in measuring the surprise latent construct (i.e., some items pairs have a local dependence between the items; Huggins-Manley & Han, 2017). Some

researchers (Cole, Ciesla, & Steiger, 2007) have argued that researchers can better specify a latent variable by adding correlated residuals for pairs of items that likely exhibit such forms of local dependence. Results of correlated residuals are available in the online supplemental materials.

Perceptions of information quality. After reading the persuasive message, participants responded to two items: "To what extent do you perceive that APL News provided good reasons for supporting the university service program?" (1 = *the reasons for supporting the university service program were very bad*, 9 = *the reasons for supporting the university service program were very good*) and "To what extent did you find APL News's reasons for supporting the university service program to be compelling?" (1 = *not at all compelling*, 9 = *very compelling*),  $r = .80$ .

Attitudes toward the university service program. Attitudes toward the university service program were measured with three items: "How much do you support the university service program?," "How much is the university service program a good idea?," and "To what extent do you think that the university service program would be beneficial?" anchored at (1 = *not at all*, 9 = *very much*),  $\alpha = .96$ .

## Results

We took an integrative data analysis approach to presenting our findings. An integrative data analysis is a way to examine multiple data sets with potentially different measures in a SEM framework (Curran & Hussong, 2009). Although the issue of different measures is less relevant to the current three studies, it was relevant to the next set of studies, and we wanted to be consistent across results presentations. The integrative data analysis approach to testing our hypotheses allowed us to control for measurement error to the best of our ability, which should help to minimize inflation of the Type I error rate when examining influences of one variable when controlling for others (Westfall & Yarkoni, 2016). We first present the experimental effects of bias and untrustworthiness on surprise to address the primary goal of the research. Next, we present a mediation model, which allowed us to test several predictions at once, including examining our manipulation checks, examining the effects of the perceived bias and trustworthiness latent variables on surprise, and testing our complete indirect paths through the inference of information quality to attitudes.

**Surprise.** To provide an experimental test of the effects of source bias and untrustworthiness on surprise, we examined a latent variable model in which we entered the bias and untrustworthiness manipulations, as well as their interaction as observed exogenous variables predicting the surprise latent variable. In this analysis, there was a significant effect of source bias on surprise,  $b = .50$ ,  $SE = .09$ ,  $z = 5.42$ ,  $p < .001$ , but not an effect of untrustworthiness,  $b = .03$ ,  $SE = .13$ ,  $z = .24$ ,  $p = .81$ , or an interaction,  $b = .04$ ,  $SE = .09$ ,  $z = .40$ ,  $p = .69$ .<sup>2</sup> This finding supported our hypothesis that source bias would have a positive effect on surprise at position switching, but source untrustworthiness would not. This model was aimed at examining only the specific experimental effects, and it did not control for other potential influences of the manipulations on related source perceptions. Perhaps as a result, the model fit was weak on some indices (though not as poor on others): root mean square error of approximation (RMSEA) = .14, 90% confidence interval (CI) [.12, .15], standardized root mean square residual (SRMR) = .079, Tucker-Lewis Index (TLI) = .83,  $\chi^2(df = 21) = 235.08$ ,  $p < .001$ . For each SEM analysis, we report the  $\chi^2$  to be consistent with SEM conven-

tions. However, as researchers have identified that the  $\chi^2$  test is sensitive to sample size (Bentler & Bonett, 1980), we rely on the RMSEA, SRMR, and TLI to interpret the fit of our models. As examined more completely in the mediation model, model fit was substantially improved (despite addressing additional outcome measures) when including the latent variables of the source perceptions in the model.

We also examined experimental effects of bias and untrustworthiness on perceptions of information quality and persuasion, but only negative effects of untrustworthiness on these variables were supported, suggesting that there might be a variable, perhaps credibility, suppressing the positive effect of surprise, which Study 3 allowed us to test in the mediation model below. The lack of total effects for bias does not remove the possibility of indirect effects through surprise (Rucker, Preacher, Tormala, & Petty, 2011; see online supplemental materials for patterns of source manipulation effects on perceptions of information quality and attitudes for Studies 1–3). To succinctly present the downstream consequences of surprise, we examined a sequential mediation model with surprise and source credibility predicting perceptions of information quality, which then predicted attitudes.

### Mediation of bias and untrustworthiness influences on attitudes by surprise through perceptions of information quality.

Using the lavaan package in R (Rosseel, 2012), we entered the bias and untrustworthiness manipulations, as well as their interaction as observed exogenous variables. We did not expect an interaction, but we wanted to allow for this possibility. Additionally, we entered perceptions of source bias, source trustworthiness, source credibility, surprise at position switching, information quality, and attitudes as latent endogenous variables. We also included each of the indicators for these latent variables as observed variables.

First, this model allowed us to examine our manipulation checks on perceptions of bias and untrustworthiness. The bias manipulation had effects on both bias and untrustworthiness, but had a larger effect on bias, highlighting the utility of examining the latent variables of the source perceptions and controlling for them if necessary. The untrustworthiness manipulation solely affected untrustworthiness.

The effects of the bias and untrustworthiness latent variables mirrored the experimental analyses above: bias had a positive effect on surprise, whereas trustworthiness was unrelated to surprise. Furthermore, the paths from surprise to perceptions of information quality to attitudes were significant (Figure 1, full model details are available in the online supplemental materials).<sup>3,4</sup> More

<sup>2</sup> The test statistic for all SEM analyses is a Wald test, which uses a  $z$  distribution.

<sup>3</sup> When we compared models in which the parameter estimates were constrained to be equal versus not across studies, there was no difference between these models,  $\chi^2$  difference ( $df = 13$ ) = 8.30,  $p = .82$ , suggesting no moderation by study. (In this model, we had to remove the credibility latent variable since it was only included in Study 3).

<sup>4</sup> Readers may also wonder about relations between bias and trustworthiness. When we run the same SEM in each of the sets of studies, but with an added residual correlation between bias and trustworthiness, we find that the two are more often related than not: Study 1–3:  $b = -.21$ ,  $p = .19$ , Study 4–6:  $b = -.39$ ,  $p = .03$ , Study 7:  $b = -.64$ ,  $p < .0001$ . Including the bias-trustworthiness residual correlation in the model does not change any of the paths of interest in the SEMs because they were already examining unique influences of the bias latent variable while controlling for influences of the trustworthiness latent variable.

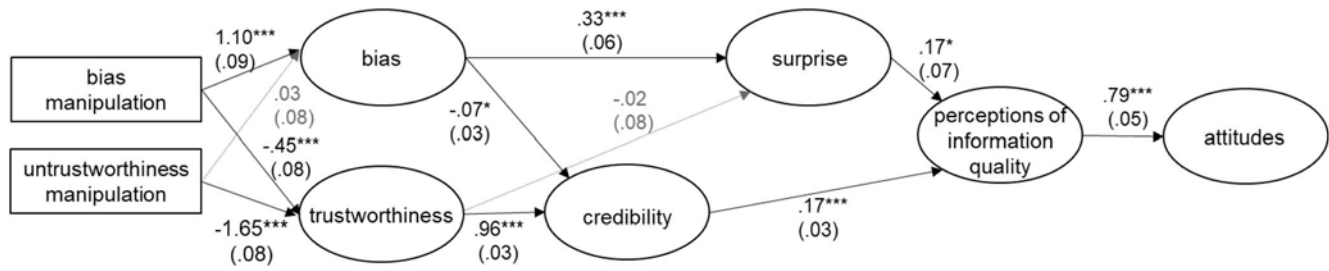


Figure 1. Mediation of effects of bias and trustworthiness through credibility and surprise to perceptions of information quality to attitudes in Studies 1–3. Latent variable indicators are omitted for presentation purposes. Latent variables are represented with ovals. Standard errors for each path are in parentheses.  $^*p < .05$ .  $^{**}p < .01$ .  $^{***}p < .001$ .

important, bootstrapping one thousand samples resulted in a significant positive indirect effect of source bias on persuasion through surprise,  $b = .047$ , 95% CI [.02, .09]. Because this indirect effect is the product of four paths, the compressed scale will inevitably make it seem small (i.e., the estimate falling relatively close to zero). Therefore, we also wanted to provide an estimate of its significance. Therefore, we examined the percentage of estimates of the indirect effect from the bootstrap samples that fell on the opposite side of zero. None of the estimates of the indirect effect (0%) crossed zero.

The paths from the bias latent variable to credibility and from credibility to perceptions of information quality were also significant, resulting in a significant negative indirect effect from the bias manipulation to attitudes through the credibility pathway,  $b = -.01$ , 95% CI [-.02, -.001]. Only 1.5% of estimates of this indirect effect from the bootstrap samples fell on the other side of zero. Thus, bias had opposite downstream effects on attitudes through the surprise and credibility pathways.

Additionally, credibility significantly mediated the negative effect of untrustworthiness on attitudes,  $b = -.21$ , 95% CI [-.28, -.12]. None of the estimates of the indirect effect from the bootstrap samples fell on the other side of zero. However, the indirect effect of untrustworthiness through surprise did not reach significance,  $b = .01$ , 95% CI [-.07, .03]. A full 39.5% of the estimates of that indirect effect from the bootstrap samples fell on the other side of zero, as the path from the trustworthiness latent variable to surprise was not significant. The model met conventional standards for acceptable fit according to our fit indices of interest: RMSEA = .069, 90% CI [.063, .076], TLI = .940, SRMR = .084,  $\chi^2(df = 305) = 936.44$ ,  $p < .001$ .

## Discussion

Combined analyses of Studies 1, 2, and 3 demonstrated that perceptions of greater source bias led people to develop stronger expectations that the source would maintain the same position. Conversely, source untrustworthiness did not have this effect, highlighting that bias and untrustworthiness, though both negative perceptions, can have different influences. Furthermore, the current data validate the notion that violating an expectation of position consistency can push toward a source being more persuasive through an inference that there are good reasons to support the new position. Thus, Studies 1–3 provide empirical support for an attributional mechanism initially described by Eagly et al. (1978)

and show that such attributions are enhanced by perceptions that the source is biased.

In the current research, surprise at position switching mediated effects of bias on persuasion through an inference of information quality, though the overall bias manipulation effects on persuasion were not significant. The inclusion of a credibility measure in Study 3 allowed us to test whether this occurred because of the negative effects that source bias can have on source credibility. Indeed, we found evidence that source bias can instigate opposing effects on persuasion through the surprise and credibility mechanisms.

However, some readers may wonder how to compare the observed lack of total effect on persuasion with Eagly et al.'s (1978) findings that persuasion increased when a source's position changed rather stayed the same in knowledge bias conditions. One key difference between the studies is that they compare different kinds of conditions. The current studies compared a biased source who switched positions and an unbiased source who switched positions. In the Eagly et al. (1978) research, the comparison was between a source that switched positions and one that did not. If we had included a condition in which a biased source continued to support their original position, we would expect the biased source to be relatively unpersuasive (as demonstrated in Wallace et al., 2019). That relatively unpersuasive comparison point could well lead to greater surprise-based persuasion when the biased source switches positions than when he or she does not.

## Studies 4, 5, and 6

One possible concern about the previous studies is that the bias manipulation might convey some information about how consistent the source is separately from the information that conveys source bias. That is, when participants read that the biased source "could only see things through their ideological lens" whereas the objective source "sometimes takes liberal positions; sometimes takes conservative positions," these statements might have communicated that the objective source has been less consistent than the biased source apart from any effect that perceived bias has on consistency. Because perceptions of consistency or surprise at inconsistency constitute the mechanism of interest, one cannot use measures of this potential "confound" to identify independent effects of perceived bias per se. Therefore, in Studies 4 and 5, we simply labeled the source

as biased or objective without any additional information about consistency or inconsistency of past positions taken. In this case, any effects of the bias manipulation on surprise must be because of a link between bias and surprise at position switching rather than because of any extraneous information that might separately convey something about position consistency. Studies 4 and 5 were identical except that in Study 4, we asked participants how surprised they would be if the source switched positions and in Study 5 we told participants that the source had switched positions and asked how surprised they were.

Although the direct manipulations of source bias used in the studies reported thus far reflect many times in life when people label others or themselves as biased, clearly there should also be times when people infer bias without being told directly that the source is biased (for an example of this, see Study 4 of Wallace et al., 2019). Therefore, in Study 6, we wanted to examine whether bias would have the same influences on surprise at position switching and indirectly on persuasion using a subtler manipulation of bias. The design of Study 6 mirrored the design of Studies 4 and 5 with the exception that the source bias manipulation required participants to naturally infer bias (i.e., without being told explicitly that the source was biased), ideally without also inferring untrustworthiness. Additionally, Study 6 contained a measure of source expertise immediately following the source bias and untrustworthiness manipulation checks because an initial study using this manipulation (reported in the online supplemental materials) demonstrated a positive effect of the bias manipulation on source credibility, suggesting the possibility that this manipulation had unintended influences on perceptions of source expertise.

Beyond addressing concerns with our previous manipulations, we sought to provide a better measure of inferences of information quality. The rationale for surprise at position switching would seem to rely to some degree on perceptions that it would require more compelling evidence to get a biased source to switch positions compared with an objective source. Yet, this aspect of the inferences was not directly measured in Studies 1–3. As an initial look at this facet of the attribution, we ran a within-subjects pilot study in which participants encountered sources described as biased, untrustworthy, objective, and honest (see online supplemental materials for full details). For each, they rated how compelling new information would have to be and how much information in support of the new position the source would need to switch positions. Participants reported that the biased compared with the objective source would need more compelling information ( $M_s = 5.98$  vs.  $5.00$ ),  $F(1, 49) = 11.59$ ,  $p = .001$ , and would need more information to switch positions ( $M_s = 6.06$  vs.  $5.18$ ),  $F(1, 49) = 11.08$ ,  $p = .002$ . In contrast, the untrustworthy source would need marginally less compelling information ( $M = 4.64$ ) than a trustworthy source ( $M = 5.28$ ) to switch positions,  $F(1, 49) = 2.94$ ,  $p = .09$ , and there was no significant difference in the amount of information needed for untrustworthy versus trustworthy sources to switch positions ( $M_s = 5.12$  vs.  $5.40$ ),  $F(1, 49) = .65$ ,  $p = .42$ . Therefore, to include this aspect of the crucial attribution in the next studies, we also included measures of how much and how compelling the information would have to be for source to switch positions.

## Method

**Participants.** There were 421 MTurk workers ( $N = 104$  in Study 4,  $N = 117$  in Study 5, and 200 in Study 6) recruited to participate in these studies. Three participants were excluded who reported that they did not take the study seriously.

**Design and procedure.** In Studies 4–6, we only manipulated perceptions of bias, and our manipulation of bias did not contain any information regarding source position consistency. In Studies 4 and 5, the source was simply described as “biased” or “objective” in their reporting of the university service program. Because we were limiting the information about the source, we shifted the cover story slightly. Rather than telling participants that MediaReports would be giving them information about APL News, we asked them to imagine that a friend had told them about APL News.

In Study 6, we used the issue of nuclear power as the key topic. Because this was a familiar attitude object (unlike the university service program), we had participants report their premessage attitudes toward expanding the use of nuclear power in Canada among several filler attitude items on topics such as legalizing marijuana in New Mexico, recycling, and exercising. Afterward, participants read about the source, John Myers, who had written an article for a national news website supporting nuclear power. They received information suggesting that he was either previously employed in the nuclear power industry or not.

After receiving the bias manipulation, participants responded to manipulation checks and additional source characteristics (including a measure of source expertise in Study 6) and surprise at position switching. In Studies 4 and 5, participants received the article from Studies 1–3 in which the source switched positions. In Study 6, participants were simply told that the source switched positions, but were not provided with a message. Afterward, participants reported their perception of the source’s information quality and their attitudes toward the topic.

### Independent variables.

**Perceptions of bias.** In Studies 4 and 5, in both the biased and objective conditions, participants were told that the source (APL News) had previously opposed the university service program. Participants read, “APL News is [biased/objective] in their reporting of the university service program,” depending on random assignment to condition. Additionally, in both conditions, they read, “APL News is also committed to being honest in their reporting of the university service program.” We counterbalanced whether they received information about the source’s honesty or bias/objectivity first.

In Study 6, in the objective condition, participants read that John Myers was a reporter for a local news publication. In the biased condition, they read that John Myers was a retired CEO of a nuclear power company, who sometimes wrote articles for a local news publication. In both conditions, they were told that John Myers had written about several different topics, including one article last year about the positives of building more nuclear power plants in Canada. We predicted that when the source was the former CEO of a nuclear power company, participants would perceive him as relatively biased but not necessarily untrustworthy (cf. Eagly et al., 1978).

### Dependent measures.

Premessage attitudes toward nuclear power. In only Study 6, premessage attitudes toward nuclear power were measured with two 7-point items that were mixed in with filler items. The items were, “To what degree do you support expanding the use of nuclear energy in Canada?” and “To what degree are you in favor of nuclear power in Canada?,” both anchored with (1 = *not at all*, 7 = *very much*),  $r = .92$ .

Perceptions of bias. Participants were asked the same two bias items from Studies 1–3. We included two additional but similar items to reduce measurement error: “To what extent do you perceive APL News as politically biased in their perception of the university service program?” and “To what extent do you perceive that APL News’s reporting reflects their ideological preferences?,” both anchored with (1 = *not at all*, 9 = *very much*). Study 6 used similar items, but they were adapted for the nuclear power context. Because the bias and untrustworthiness items were slightly different for each study, we calculated reliabilities separately for Studies 4 and 5 and 6, Studies 4 and 5:  $\alpha = .88$ , Study 6:  $\alpha = .94$ .

Perceptions of trustworthiness. We also measured perceptions of trustworthiness with four items to use in the SEM analyses to control for any influences of the bias manipulation on perceived untrustworthiness and to enable us to examine relations between perceived untrustworthiness and surprise at position switching. In Studies 4 and 5, participants responded to: “How much do you see APL News as truthfully communicating the facts as they see them?,” “How much do you perceive APL News as committed to being honest?,” “To what extent do you perceive that APL News reporters try to convey the truth as they see it?,” and “To what extent do you perceive that APL News reporters are willing to manipulate their readers?” (1 = *not at all*, 9 = *very much*). When we examined correlations between indicators *before* analyses, we discovered that this fourth item was not very highly correlated with the other items, ( $r = .25-.35$ ), so we did not include it as an indicator of the source trustworthiness latent variable in the SEM analyses. In Study 6, the items were very similar except that they referred to John Myers and the fourth item was replaced with “To what extent does it seem like John Myers wants to be honest about his perspective?,” Studies 4 and 5:  $\alpha = .90$ , Study 6:  $\alpha = .94$ .

Perceptions of expertise. In only Study 6, perceptions of expertise were measured with four items on 9-point scales (e.g., “To what extent does it seem like John Myers is an expert on nuclear power in Canada?”; 1 = *very inexpert*, 9 = *very expert*),  $\alpha = .94$ .

Perceptions of credibility. Credibility was measured the same as in Study 3 except that it contained an extra item measuring the extent to which participants viewed the source as a “high quality source of information,”  $\alpha = .94$ .

Surprise at position switching. In Studies 5 and 6, the surprise measure was identical to the one in Study 1. It was very similar in Study 4, but asked participants how surprised they would be if the source switched rather than telling them the source had switched,  $\alpha = .89$ .

Perceptions of information quality. In Studies 4 and 5, the first two items used to measure perceptions of information quality were identical to Studies 1–3. In Studies 4 and 5, we also asked participants “How compelling did the information supporting the university service program have to be in order for APL News to provide an article supporting the university service program?” and

“How much information in support of the university service program did APL News reporters have to learn to provide an article supporting the university service program?” The items were similar in Study 6 but referred to John Myers and nuclear power instead of APL News and the university service program,  $\alpha = .84$ . Similar to the approach we took to the pairs of items in the surprise index, in each of our analyses, we allowed the residual variances of each of the pairs of items making up the perceptions of information quality index to correlate with one another. The first pair of items included the two judgments about the quality of the arguments provided. The second pair of items included the judgments about how much and how compelling information supporting the new side would have to be for the source to switch. Results of correlated residuals are available in the online supplemental materials.

Attitudes. Attitudes toward the university service program/nuclear power were measured the same as in Studies 1–3,  $\alpha = .97$ .

## Results

Integrative data analysis requires the specification of a “common item” for each latent variable in the model. Ideally, this would be an item that is shared across each dataset. In the previous set of analyses, the measures were identical across studies so this was not an issue. In the current set of analyses, all the latent variables shared at least one common item across studies with some variation in the wording of the bias and trustworthiness items across samples, largely because of shifts in topics across studies and the biases inherent to those situations—for example, ideological versus personal. The online supplemental materials contains details of the common items we used for this integrative analysis, which are identical to the common items used across integrative analyses in each section of the article.

**Surprise.** To provide an experimental test of the effects of source bias on surprise, we examined a latent variable model in which we entered the bias manipulation as an observed exogenous variable predicting the surprise latent variable. In this analysis, there was a significant effect of source bias on surprise,  $b = .29$ ,  $SE = .07$ ,  $z = 4.00$ ,  $p < .001$ , once again supporting our hypothesis that source bias would have a positive effect on surprise at position switching. Though not the key point of this analysis, the model fit was as follows: RMSEA = .08, 90% CI [.05, .11], SRMR = .026, TLI = .96,  $\chi^2(df = 11) = 39.61$ ,  $p < .001$ . We once again did not observe significant total effects of the source bias manipulation on perceptions of information quality or attitudes (results reported in the online supplemental materials).

**Mediation of bias influences on attitudes by surprise and credibility through perceptions of information quality.** As shown in Figure 2, we conducted the same SEM analyses as in Studies 1–3 except that Studies 4–6 did not contain an untrustworthiness manipulation and did contain a measure of source expertise in Study 6. Paralleling previous studies, the bias manipulation had a significant effect on perceived source bias. Bias then affected both surprise and perceived source credibility, which each affected perceptions of information quality, which in turn affected people’s attitudes. Ultimately, source bias had a positive indirect effect on attitudes through surprise,  $b = .030$ , 95% CI [.011, .051], 0.1% of estimates of the indirect effect from bootstrap samples crossed zero, but a negative indirect effect through credibility,



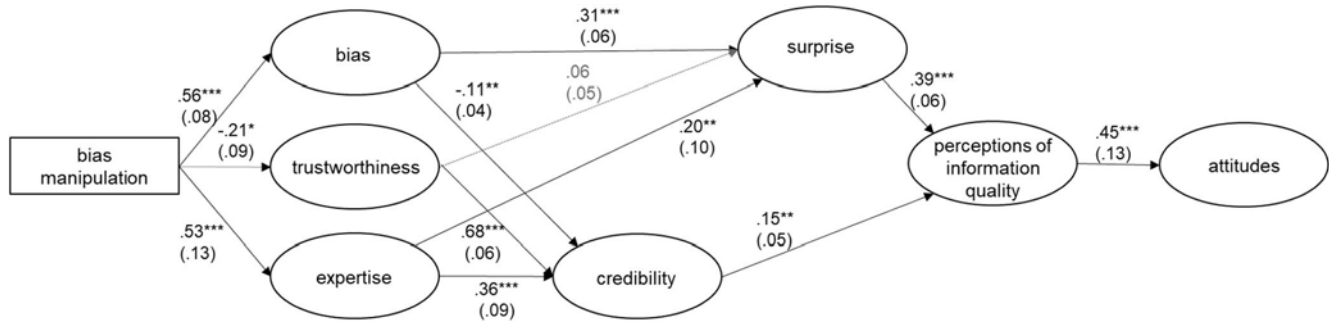


Figure 2. Mediation of effects of bias through surprise and credibility to perceptions of information quality to attitudes in Studies 4–6. Latent variable indicators are omitted for presentation purposes. Latent variables are represented with ovals. Standard errors for each path are in parentheses. <sup>\*\*\*</sup> $p < .05$ . <sup>\*\*</sup> $p < .01$ . <sup>\*</sup> $p < .001$ .

$b = -.004$ , 95% CI  $[-.009, -.001]$ , 0.5% of estimates of the indirect effect from bootstrap samples fell on the other side of zero.<sup>5</sup> The fit indices were as follows: RMSEA = .053, 90% CI  $[.049, .058]$ , SRMR = .13.<sup>6</sup> The model  $\chi^2(df = 448) = 979.14$ ,  $p < .001$ .

Thus, even without any additional information that might communicate something about the source's consistency separate from the source's bias, we still found that our manipulations of source bias influenced participants' surprise at position switching. This demonstrates that the perception of bias itself can create expectations that the source will maintain their original position. Additionally, Study 6 demonstrated that these effects also occur when participants infer source bias rather than being told directly that the source is biased or objective. Furthermore, they can even occur when the source does not provide a message for their new position, but when participants simply know that the source has switched positions.

In summary, these studies provide additional evidence for the unique positive effect of source bias on surprise, which results in a positive indirect effect on persuasion. Once again, perceptions of source untrustworthiness did not affect surprise, highlighting that these different source perceptions can have differing consequences.

### Study 7: Preregistered Replication

Although the previous studies represent the complete set of studies that we had collected before this replication study and the results across studies are quite consistent, readers may have concerns about individual study power. Additionally, because we had not preregistered an analysis plan for the previous studies, readers might have concerns about potential flexibility in analysis choices. To address these concerns and provide a final test of our hypotheses, we conducted a highly powered preregistered replication study (preregistration link: [https://osf.io/m38vk/?view\\_only=048d25c77c3946a1915797885637c9a4](https://osf.io/m38vk/?view_only=048d25c77c3946a1915797885637c9a4)).

### Method

**Participants.** We determined our target sample size by conducting power analyses using Monte Carlo simulations of our planned SEM analyses with conservative estimates of our effect sizes. For each path, the estimate we used was smaller than the

estimate obtained based on an integrative data analysis of all of the data we had collected up to this point (reported in the integrative data analysis section below). The paths from the bias manipulation to the perceived bias latent variable and from perceptions of information quality to attitudes were comparatively larger so we used an unstandardized coefficient of .40 in our power analyses for those paths. The paths from perceived bias to surprise and from surprise to perceptions of information quality were comparatively smaller so we used an unstandardized coefficient of .15 for each of these paths in our power analyses. The model is specified in terms of those direct paths, but these path values imply an estimate of the indirect effect of .0036 (based on a multiplication of the direct path

<sup>5</sup> To test whether these results differed across studies, we compared models in which the parameter estimates were constrained to be equal versus not. Because these studies used some different measures, we had to remove the expertise latent variable from the model and use a composite of the bias and untrustworthy measures rather than estimating their latent variables. We found that these models significantly differed,  $\chi^2$  difference ( $df = 12$ ) = 38.71,  $p < .001$ . The only primary path of interest that differed was the effect of information quality on attitudes, which was smaller in Study 6, likely reflecting that attitudes were harder to change in that study because people already formed attitudes toward nuclear power before the study,  $\chi^2$  difference = 11.70,  $p = .003$ . This path was directionally consistent across studies. Chi-square difference tests for all paths in this model are reported in the online supplemental materials.

<sup>6</sup> TLI is not available for this model because of the nonoverlapping measures across studies (e.g., expertise only in Study 6). This relatively high SRMR value seems to be driven by pockets of high residuals between final attitudes and perceptions of source credibility, trustworthiness, and bias and between measures of trustworthiness and expertise in Study 6 that do not occur in the other studies (see online supplemental materials for a table of Study 6 residuals, as well as individual study fit indices). We examined whether including paths connecting these source perception latent variables to the final attitude latent variable would decrease SRMR. This model continued to have good fit according to RMSEA, but SRMR was unchanged. More importantly, including these paths did not change any of the other mediational paths (only slightly strengthening the link between perceived information quality and final attitudes). Although we cannot fully diagnose the relatively high SRMR values for Study 6 (or the integrative data analyses including that study), the parameter estimates from Study 6 continue to support the key mediational paths, and the study's inclusion with other data does not influence the RMSEA values (only the SRMR values). When we conduct an IDA with just Studies 4 and 5, it demonstrates acceptable fit for both RMSEA = .063, 90% CI  $[.05, .07]$  and SRMR = .063, and the significance of the paths of interest do not change. The results of that IDA are available in the online supplemental materials.

values—i.e., .4<sup>[.15, .15]</sup>. For comparison, standardized values of the same direct paths would be .21 for the path from the bias manipulation to the perceived bias latent variable, .20 for the perceived bias to surprise path, .21 for the surprise to information quality path, and .4 for the information quality to attitudes path. These standardized values for the direct paths would imply an indirect effect very similar in value to the unstandardized values (.0035 = .21<sup>[.15, .15]</sup>.4). Monte Carlo analyses using these values as population values provide estimates of power to detect effects of these sizes at over 98% for each of the direct paths and at 86% for the indirect effect based on a sample of 600 participants. Estimates of power for each path are available in the preregistration document. Our power calculations were based on those conservative estimates, so the closer population values come to the average values of the previous studies, the higher the actual power would be. Requesting 600 participants from TurkPrime resulted in 603 complete lines of data. Consistent with our preregistered exclusion criteria, we excluded 22 participants for reporting a “5” on our attention check and 28 participants for providing nonsense or two incorrect answers to our Winograd questions, leaving 555 participants for analyses.<sup>7</sup> Also consistent with our preregistration, we excluded 69 responses that were greater than 3 *SDs* from the mean on an item used in our primary analyses. Use of the remaining items assessing that variable did not reduce the number of participants in the analysis, and for a given item, no more than 15 responses were omitted. To handle this missing data, in our SEM analyses, we used a maximum likelihood estimator based on the available responses.

**Design and procedure.** This study used a 2 (Bias: high vs. low) × 2 (Untrustworthiness: high vs. low) design. Other than this addition of an untrustworthiness manipulation, the study was almost identical to Studies 4 and 5. As in those studies, the topic was the university service policy.

#### Independent variables.

Perceptions of bias and untrustworthiness. Depending on random assignment to condition, participants read that, “APL News is known to be politically [biased/objective], and also [dishonest/honest] in their reporting.” Whether the bias or untrustworthiness information came first was counterbalanced.

**Dependent measures.** The bias ( $\alpha = .89$ ), trustworthiness ( $\alpha = .95$ ), credibility ( $\alpha = .97$ ), surprise ( $\alpha = .87$ ), perceptions of information quality ( $\alpha = .85$ ), and attitude ( $\alpha = .96$ ) measures were almost identical to Studies 4 and 5. The one exception was that we replaced the fourth trustworthiness item with, “How much do you perceive that APL News wants to be honest? (1 = *not at all*, 9 = *very much*).”

## Results

**Surprise.** Once again, we began by experimentally testing the effects of bias and untrustworthiness on surprise. In a SEM analysis, we entered the bias and untrustworthiness manipulations along with their interaction as observed exogenous variables predicting surprise, which was entered as a latent endogenous variable with each of its indicators. Replicating our previous results, there was a significant positive effect of bias on surprise,  $b = .24$ ,  $SE = .07$ ,  $z = 3.32$ ,  $p = .001$ . Conversely, there was no effect of untrustworthiness,  $b = .04$ ,  $SE = .07$ ,  $z = .56$ ,  $p = .58$ , and no interaction,  $b = .04$ ,  $SE = .07$ ,  $z = .59$ ,  $p = .56$ . The model fit was

as follows: RMSEA = .08, 90% CI [.06, .09], SRMR = .04, TLI = .94,  $\chi^2(df = 21) = 89.70$ ,  $p < .001$ . Therefore, this replication study provided additional evidence for the differing effects of bias and untrustworthiness on surprise.

**Mediation of bias and untrustworthiness influences on attitudes by surprise and credibility through perceptions of information quality.** Next, we conducted SEM analyses paralleling those in Studies 1–3.<sup>8</sup> As shown in Figure 3, the results are very consistent with previous results and clearly support the proposed indirect effect of source bias through surprise at position switching. Various fit indices indicated acceptable fit according to Hu and Bentler’s (1999) recommendations, RMSEA = .06, 90% CI [.06, .07], TLI = .94, and SRMR = .07, so we did not consider model modifications. Additionally,  $\chi^2(df = 306) = 969.22$ ,  $p < .001$ .

In terms of the structural model, the first thing to note is that, similar to the previous studies, the bias manipulation had a significant effect on the bias latent variable and a smaller but significant effect on the trustworthiness latent variable. The untrustworthiness manipulation had a significant effect on the trustworthiness latent variable and a smaller but significant effect on perceptions of bias. The bias latent variable positively predicted surprise at position-switching, whereas the trustworthiness latent variable was not significantly related to surprise. Additionally, the bias latent variable negatively predicted perceptions of source credibility whereas the trustworthiness latent variable positively predicted perceptions of source credibility. The more participants initially perceived APL News as credible and, independently, the more they were surprised when the source switched positions, the more they inferred that the source had good reasons for the new position. Inferring that the source had good reasons for the position ultimately led to more favorable attitudes toward the university service policy.

Ultimately, this resulted in a significant positive indirect effect of source bias on attitudes through surprise,  $b = .07$ , 95% CI [.04, .11] with 0.0% of the estimates of the indirect effect from the bootstrapped samples crossing zero. However, as in the previous study sets, there was also a negative indirect effect of source bias through credibility,  $b = -.01$ , 95% CI [-.01, -.001], with 0.09% of the estimates of the indirect effect from the bootstrapped sam-

<sup>7</sup> This resulted in dropping more participants than we had expected. However, even with only 555 participants, we still have 82% power to detect the indirect effect of interest and over 98% power to detect each direct path of interest, based on the conservative effect size estimates we used to determine our preregistered sample size.

<sup>8</sup> Based on reviews expressing a preference for bootstrapping of standard errors, we agreed to shift our initial analysis plans to bootstrap standard errors instead of bootstrapping the estimates themselves. When preregistering the analysis, we inadvertently added the bootstrapping of standard errors to the code that bootstrapped the estimates, which is not possible. The two options are to: (a) bootstrap the standard errors and obtain the indirect effect estimate from those samples, or (b) bootstrap the model after obtaining standard errors based on the delta method to obtain the indirect effect estimate. In the text, we present analyses based on bootstrapped standard errors. However, the results are almost identical when we conduct analyses based on the second method. These results, along with the R code we used for each are available in the online supplemental materials. To be consistent, we also report Integrative Data Analyses based on the bootstrapped standard errors, but results based on the delta method (and then bootstrapped to test the indirect effect) are also available in the online supplemental materials.

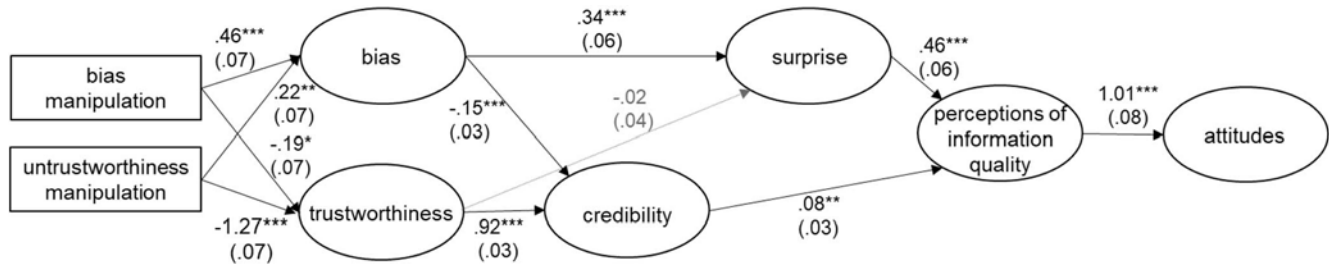


Figure 3. Mediation of effects of bias and trustworthiness through credibility and surprise to perceptions of information quality to attitudes in Study 7. Latent variable indicators are omitted for presentation purposes. Latent variables are represented with ovals. Standard errors for each path are in parentheses.  $^*p < .05$ .  $^{**}p < .01$ .  $^{***}p < .001$ .

ples crossing zero. Additionally, as in the previous study sets, there was a significant indirect effect of untrustworthiness on persuasion through credibility,  $b = -.09$ , 95% CI  $[-.15, -.03]$ , with 0.09% of the estimates of the indirect effect from the bootstrapped samples crossing zero. Because the path from trustworthiness to surprise was nonsignificant, there was not a significant indirect effect of untrustworthiness through surprise,  $b = .01$ , 95% CI  $[-.03, .05]$ , with 32% of the estimates of the indirect effects from the bootstrapped samples crossing zero.

In summary, this highly powered preregistered replication provided additional support for the primary hypotheses. When sources switched positions, source bias had a positive effect on surprise at the switch. Conversely, untrustworthiness did not significantly affect surprise. Further, surprise had a positive downstream effect on how much participants perceived that the source had compelling information in support for the new position, which in turn increased favorability toward the policy. These results highlight the necessity of considering source bias and untrustworthiness as conceptually distinct perceptions with differing consequences.

### Integrative Data Analysis of Complete Set of Studies

As a final analysis, we conducted an integrative data analysis of all studies in this package (that comprise all complete studies that we can use to test the full mediation model; Table 1). This analysis allowed us to provide an average effect size estimate and examine the strength of evidence for our effects across all the data we have

to test them. Additionally, we compared integrative analyses with and without the replication study (see Table 2), which allowed us to examine the effect of the preregistered replication study on the evidence for our effects.

When we conducted the integrative data analysis with all of the data, all primary paths of interest (bias manipulation to perceived source bias, perceived bias to surprise, surprise to perceptions of information quality, and perceptions of information quality to attitudes) were highly significant,  $z_s > 8.50$ ,  $p_s < .0001$  (see Table 1), suggesting that across studies there was strong evidence for every path of interest. Further, as is evidenced by the change in  $z$ -values when adding the preregistered replication, these paths became somewhat more significant than in the integrative analysis without the replication study, so the addition of the replication study increased confidence that the previously estimated effects were correct in the direction of their difference from zero. The effect size for the relation between perceived source bias and surprise was the same in both integrative analyses, so our best estimate of the effect size for that path did not change with the addition of the replication study. However, this integrative analysis demonstrated larger effect sizes for the paths from surprise to perceptions of information quality and from information quality to attitudes, suggesting that in the initial studies, we may have been underestimating the strength of these paths.

Further, the indirect effect of bias through surprise,  $b = .06$ , 95% CI  $[.04, .08]$ , 0.0% of 10,000 estimates from bootstrapped samples crossing zero, was supported as well. Overall, this effect

Table 1

Integrative Data Analysis of Complete Set of Studies

Path	Coef.	SE	$z$	$p$
Bias manipulation to perceived bias	.73	.05	15.13	<.0001
Bias manipulation to perceived trustworthiness	-.31	.06	-5.29	<.0001
Perceived bias to surprise	.32	.04	8.92	<.0001
Perceived trustworthiness to surprise	.03	.03	1.02	.308
Perceived bias to credibility	-.11	.02	-5.67	<.0001
Perceived trustworthiness to credibility	.88	.02	49.86	<.0001
Surprise to perceived information quality	.29	.03	9.59	<.0001
Credibility to perceived information quality	.05	.02	2.39	.017
Perceived information quality to attitudes	.95	.05	18.09	<.0001
Bias manipulation to perceived expertise	.63	.12	5.25	<.0001
Perceived expertise to surprise	.20	.08	2.50	.012
Perceived expertise to credibility	.17	.06	2.99	.003

Table 2

Integrative Data Analysis Without Preregistered Replication

Path	Coef.	SE	$z$	$p$
Bias manipulation to perceived bias	.86	.06	13.84	<.0001
Bias manipulation to perceived trustworthiness	-.37	.07	-5.07	<.0001
Perceived bias to surprise	.32	.04	7.40	<.0001
Perceived trustworthiness to surprise	.09	.04	2.41	.016
Perceived bias to credibility	-.08	.03	-3.26	.001
Perceived trustworthiness to credibility	.86	.03	31.62	<.0001
Surprise to perceived information quality	.20	.04	4.63	<.0001
Credibility to perceived information quality	.17	.03	5.65	<.0001
Perceived information quality to attitudes	.60	.07	8.40	<.0001
Bias manipulation to perceived expertise	.61	.12	5.01	<.0001
Perceived expertise to surprise	.26	.09	2.82	.005
Perceived expertise to credibility	.22	.06	3.44	.001

was larger than in the integrative analysis without the replication study ( $b = .03$ , 95% CI [.02, .05], 0.0% of 10,000 estimates from bootstrapped samples crossing zero), suggesting that this indirect effect might on average be bigger than we had previously observed. Additionally, the fact that even with this added data, *zero* of the bootstrapped samples crossed zero boosted our confidence in this indirect effect. The negative indirect effect of bias through credibility was also supported,  $b = -.004$ , 95% CI [-.007, -.001], 0.54% of 10,000 estimates from bootstrapped samples crossing zero. However, relative to the integrative analysis of data before the preregistered replication ( $b = -.01$ , 95% CI [-.01, -.002], 0.05% of 10,000 estimates from bootstrapped samples crossing zero), this effect size was smaller and less significant. Thus, when examining all the data available to test these hypotheses, the dual effects of bias on persuasion through surprise and credibility are clearly supported and more so than they were based on the initial six studies. In this integrative analysis, the path from trustworthiness to surprise was not supported, suggesting that this effect is likely nonexistent, inconsistent, too small to be detected with the current sample size, or subject to contextual factors that we have not yet identified. Finally, this model demonstrated acceptable and identical fit in both analyses that included or excluded the replication study, RMSEA = .047, 90% CI [.044, .050], SRMR = .17.<sup>9</sup> For the model with the replication study,  $\chi^2(df = 448) = 1921.83$ ,  $p < .001$ . Without the replication study,  $\chi^2(df = 448) = 1398.10$ ,  $p < .001$ .

### General Discussion

The current research demonstrated that perceiving a source as biased can have directionally different effects than perceiving a source as untrustworthy, despite them being similarly negative perceptions. That is, source bias leads people to expect that the source will continue to take the same position whereas source untrustworthiness does not. Expecting the source to take a particular position led participants to be more persuaded when the source took the opposite position because of an inference of high information quality in support of the new position (cf. Eagly et al., 1978). This led to a positive indirect effect of bias through surprise to make the source relatively more persuasive than an unbiased source who also switches positions when controlling for perceptions of source credibility. Previous research has demonstrated that when sources do not shift positions, source bias has led to negative influences on persuasion (Wallace et al., 2019). Thus, taken together, these lines of research have demonstrated that bias can have different influences on persuasion depending on whether the source switches positions. In comparison, untrustworthiness and low expertise have only produced negative main effects on persuasion regardless of whether the source switches positions (excluding amount of processing effects), highlighting the need to understand bias as separable from trustworthiness or expertise. Importantly, in Study 6, we replicated previous research showing that people can naturally infer bias without necessarily inferring untrustworthiness (Wallace et al., 2019), and these inferences have downstream consequences that parallel those of more direct manipulations of source bias and trustworthiness.

### Implications

**Bias and untrustworthiness as distinct perceptions in additional domains.** Persuasion represents only one domain in which the distinction between bias and untrustworthiness might be relevant. For example, traditional models of person perception have focused on perceptions of competence and warmth (with warmth sometimes divided into morality and sociability; e.g., Cuddy et al., 2008; Leach, Ellemers, & Barreto, 2007). The items typically used to capture morality tap directly into perceptions of trustworthiness (e.g., honest, sincere), but they do not generally include perceptions of bias. Indeed, it seems that none of the dimensions directly include the perception that a person is biased per se. However, this perception might be particularly important in an intergroup domain in which individuals would be concerned that others would have biases against their group in addition to being untrustworthy. In the persuasion domain, perceptions of bias can be separated from liking, trustworthiness, and expertise (competence) of message sources (see Wallace, 2015). If such perceptions are also separable in impression formation more generally, it may be important for person perception research to consider bias separately from morality (trustworthiness), sociability, or competence. Indeed, although previous research has focused on trustworthiness as a primary dimension of person perception, with people consistently preferring trustworthy over untrustworthy others, it is possible that this preference depends on the bias that the target possesses. For example, Melnikoff and Bailey (2018) recently found that people prefer honesty in spies that are working against them, but dishonesty in spies that are working for them. This general idea may be true across situations in which people perceive that others are biased in their favor versus objective or biased against them.

**Consequences of bias and untrustworthiness on other flip-flopping-related outcomes.** A good deal of research has addressed effects of observing a person contradict what they have previously said or done. Researchers have identified when this inconsistency results in perceived hypocrisy (Barden, Rucker, & Petty, 2005; Barden, Rucker, Petty, & Rios, 2014; Effron & Miller, 2015; Kreps, Laurin, & Merritt, 2017), enhanced memory for the source's position (Putnam, Wahlheim, & Jacoby, 2014), or a dissonant state that changes the observer's own attitudes and behavior (Focella, Stone, Fernandez, Cooper, & Hogg, 2016). Each of these effects might be influenced by whether the source of the inconsistency is viewed as biased or untrustworthy before the contradictory statements or behavior. For example, even though moral sources are expected to be consistent in their positions the same way that biased sources are, Kreps et al. (2017) found that, when moral leaders flip flop, they are viewed as hypocritical and, thus, are viewed as less effective leaders and garner less support. It is possible that knowing that a moral leader is also biased in their opinion before position switching could remove or reverse these

<sup>9</sup> As noted in Footnote 5, this relatively high SRMR value was driven by inclusion of Study 6. Examining an IDA using just Studies 1–5, RMSEA = .05, 90% CI [.05, .06], SRMR = .09, and an IDA using Studies 1–5 and 7, RMSEA = .06, 90% CI [.05, .06], SRMR = .08 (both IDAs available in the online supplemental materials), demonstrates acceptable for both fit indices in both analyses. Also, the key paths of interest are very similar to those in the IDA that included Study 6 (presented in Tables 1 and 2).

effects. Additionally, perceiving a target as biased or untrustworthy may influence the recipient's memory for the source's position. For example, better memory for flip flopping could develop for a biased rather than untrustworthy target, as the switch in position should be less expected for the biased target. Finally, differences in perceiving a target as biased versus untrustworthy may have downstream consequences for vicarious dissonance (Focella et al., 2016). If recipients perceive that a target is biased before behaving inconsistently, they may be less likely to experience vicarious dissonance to the extent that they perceive that the person's contradictory actions occurred as the result of learning compelling evidence in favor of the new position. Conversely, perceiving a target as untrustworthy may result in the previously observed vicarious dissonance effects, as perceived dishonesty has played a key theoretical role in self-oriented forms of dissonance theory (e.g., Aronson, 1969; Stone & Cooper, 2001).

**Biased versus fake news in the wild.** Finally, this work has practical implications: it suggests that whether people view news sources as biased or fake can have different effects when those sources switch positions. Thus, the distinction between biased and fake news not only serves to distinguish between reasons that a source might provide inaccurate information; it suggests that these two perceptions can, at times, have different consequences (perhaps even opposing consequences at times). If news sources and politicians wish to be believed when they switch positions, the current work suggests that it is essential that they maintain a perception of honesty (that is likely most easily maintained by actually being honest). The current work also suggests that, at least in position-switching settings, perceptions of source bias might not be harmful and could, in certain respects, enhance effectiveness of a message supporting a new position.

## Limitations

**Generalizability.** Each of our studies used a convenience sample from the United States that was likely to be more White, Educated, Industrialized, Rich, and Democratic than either the nation as a whole or certainly the broader world (Henrich, Heine, & Norenzayan, 2010). We do not have reasons to expect that bias and untrustworthiness should necessarily be more or less conceptually overlapping in other kinds of samples, but we would not want to draw conclusions about the effects of our specific bias and untrustworthiness manipulations among populations not represented in our sample. Consistent with the use of different operations across studies, however, the purpose of the research was not to make claims about particular operations but, instead about relations among the conceptual variables. Even there, for some parts of our research, it seems quite possible that there would be cultural limits. For example, effects of surprise (based in expectancies and expectancy violation) might be quite different in cultures that embrace dialectical thinking or that are more accepting of individuals acting differently across contexts. Such questions constitute valuable avenues for future research.

## References

Aronson, E. (1969). The theory of cognitive dissonance: A current perspective. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (4th ed., pp. 1–34). New York, NY: Academic Press. [http://dx.doi.org/10.1016/S0065-2601\(08\)60075-1](http://dx.doi.org/10.1016/S0065-2601(08)60075-1)

- Baker, S. M., & Petty, R. E. (1994). Majority and minority influence: Source-position imbalance as a determinant of message scrutiny. *Journal of Personality and Social Psychology*, *67*, 5–19. <http://dx.doi.org/10.1037/0022-3514.67.1.5>
- Barden, J., Rucker, D. D., & Petty, R. E. (2005). "Saying one thing and doing another": Examining the impact of event order on hypocrisy judgments of others. *Personality and Social Psychology Bulletin*, *31*, 1463–1474.
- Barden, J., Rucker, D. D., Petty, R. E., & Rios, K. (2014). Order of actions mitigates hypocrisy judgments for ingroup more than outgroup members. *Group Processes & Intergroup Relations*, *17*, 590–601. <http://dx.doi.org/10.1177/1368430213510192>
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588–606. <http://dx.doi.org/10.1037/0033-2909.88.3.588>
- Cole, D. A., Ciesla, J. A., & Steiger, J. H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, *12*, 381–398. <http://dx.doi.org/10.1037/1082-989X.12.4.381>
- Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS Map. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 40, pp. 61–149). San Diego, CA: Elsevier Academic Press. [http://dx.doi.org/10.1016/S0065-2601\(07\)00002-0](http://dx.doi.org/10.1016/S0065-2601(07)00002-0)
- Curran, P. J., & Hussong, A. M. (2009). Integrative data analysis: The simultaneous analysis of multiple data sets. *Psychological Methods*, *14*, 81–100. <http://dx.doi.org/10.1037/a0015914>
- Eagly, A. H., Wood, W., & Chaiken, S. (1978). Causal inferences about communicators and their effect on opinion change. *Journal of Personality and Social Psychology*, *36*, 424–435. <http://dx.doi.org/10.1037/0022-3514.36.4.424>
- Effron, D. A., & Miller, D. T. (2015). Do as I say, not as I've done: Suffering for a misdeed reduces the hypocrisy of advising others against it. *Organizational Behavior and Human Decision Processes*, *131*, 16–32. <http://dx.doi.org/10.1016/j.obhdp.2015.07.004>
- Focella, E. S., Stone, J., Fernandez, N. C., Cooper, J., & Hogg, M. A. (2016). Vicarious hypocrisy: Bolstering attitudes and taking action after exposure to a hypocritical ingroup member. *Journal of Experimental Social Psychology*, *62*, 89–102. <http://dx.doi.org/10.1016/j.jesp.2015.09.014>
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–83. <http://dx.doi.org/10.1017/S0140525X0999152X>
- Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion: Psychological studies of opinion change*. New Haven, CT: Yale University Press.
- Hovland, C. I., & Mandell, W. (1952). An experimental comparison of conclusion-drawing by the communicator and by the audience. *Journal of Abnormal and Social Psychology*, *47*, 581–588. <http://dx.doi.org/10.1037/h0059833>
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55. <http://dx.doi.org/10.1080/10705519909540118>
- Huggins-Manley, A. C., & Han, H. (2017). Assessing the sensitivity of weighted least squares model fit indices to local dependence in item response theory models. *Structural Equation Modeling*, *24*, 331–340. <http://dx.doi.org/10.1080/10705511.2016.1247355>
- Kennedy, K. A., & Pronin, E. (2008). When disagreement gets ugly: Perceptions of bias and the escalation of conflict. *Personality and Social Psychology Bulletin*, *34*, 833–848. <http://dx.doi.org/10.1177/0146167208315158>

- Kreps, T. A., Laurin, K., & Merritt, A. C. (2017). Hypocritical flip-flop, or courageous evolution? When leaders change their moral minds. *Journal of Personality and Social Psychology, 113*, 730–752. <http://dx.doi.org/10.1037/pspi0000103>
- Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology, 93*, 234–249. <http://dx.doi.org/10.1037/0022-3514.93.2.234>
- Melnikoff, D. E., & Bailey, A. H. (2018). Preferences for moral vs. immoral traits in others are conditional. *Proceedings of the National Academy of Sciences of the United States of America, 115*, E592–E600. <http://dx.doi.org/10.1073/pnas.1714945115>
- Pronin, E., Gilovich, T., & Ross, L. (2004). Objectivity in the eye of the beholder: Divergent perceptions of bias in self versus others. *Psychological Review, 111*, 781–799. <http://dx.doi.org/10.1037/0033-295X.111.3.781>
- Putnam, A. L., Wahlheim, C. N., & Jacoby, L. L. (2014). Memory for flip-flopping: Detection and recollection of political contradictions. *Memory & Cognition, 42*, 1198–1210. <http://dx.doi.org/10.3758/s13421-014-0419-9>
- Ross, L., & Ward, A. (1996). Naive realism in everyday life: Implications for social conflict and misunderstanding. In E. S. Reed, E. Turiel, & T. Brown (Eds.), *Values and knowledge* (pp. 103–135). Hillsdale, NJ: Erlbaum, Inc.
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software, 48*, 1–36. <http://dx.doi.org/10.18637/jss.v048.i02>
- Rucker, D. D., Preacher, K. J., Tormala, Z. L., & Petty, R. E. (2011). Mediation analysis in social psychology: Current practices and new recommendations. *Social and Personality Psychology Compass, 5*, 359–371. <http://dx.doi.org/10.1111/j.1751-9004.2011.00355.x>
- Stone, J., & Cooper, J. (2001). A self-standards model of cognitive dissonance. *Journal of Experimental Social Psychology, 37*, 228–243. <http://dx.doi.org/10.1006/jesp.2000.1446>
- Wallace, L. E. (2015). *Antecedents and consequences of perceiving a source as biased* (Doctoral dissertation). Retrieved from [http://rave.ohiolink.edu/etdc/view?acc\\_num=osu1430996325](http://rave.ohiolink.edu/etdc/view?acc_num=osu1430996325)
- Wallace, L. E. (2019). *Distinguishing perceptions of bias from perceptions of untrustworthiness: Independent perceptions with shared as well as unshared consequences and antecedents* (Doctoral dissertation). Retrieved from [http://rave.ohiolink.edu/etdc/view?acc\\_num=osu1557140210683552](http://rave.ohiolink.edu/etdc/view?acc_num=osu1557140210683552)
- Wallace, L. E., Wegener, D. T., & Petty, R. E. (2019). When sources honestly provide their biased opinion: Bias as distinct perception with independent effects on credibility and persuasion. *Personality and Social Psychology Bulletin*. Advance online publication. <http://dx.doi.org/10.1177/0146167219858654>
- Walster, E., Aronson, E., & Abrahams, D. (1966). On increasing the persuasiveness of a low prestige communicator. *Journal of Experimental Social Psychology, 2*, 325–342. [http://dx.doi.org/10.1016/0022-1031\(66\)90026-6](http://dx.doi.org/10.1016/0022-1031(66)90026-6)
- Westfall, J., & Yarkoni, T. (2016). Statistically controlling for confounding constructs is harder than you think. *PLoS ONE, 11*, e0152719. <http://dx.doi.org/10.1371/journal.pone.0152719>

Received March 26, 2019

Revision received October 10, 2019

Accepted October 10, 2019 